AD_____

AWARD NUMBER: W81XWH-04-1-0475

TITLE: Development of a Computer-aided Diagnosis System for Early Detection of Masses Using Retrospectively Detected Cancers on Prior Mammograms

PRINCIPAL INVESTIGATOR: Jun Wei, Ph.D.

CONTRACTING ORGANIZATION: University of Michigan
Ann Arbor, MI 48109

REPORT DATE: June 2009

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
1 June 2009	Final	1 Jun 2004 – 31 May 2009

4. TITLE AND SUBTITLE

Development of a Computer-aided Diagnosis System for Early Detection of Masses Using Retrospectively Detected Cancers on Prior Mammograms

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0475

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Wendy M. Kohrt, Ph.D.; L. Michael Glode, M.D.; Robert S. Schwartz, M.D.; Daniel W. Barry, M.D.

E-Mail: jvwei@umich.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Michigan
Ann Arbor, MI 48109

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The performance of a CAD system for subtle lesions is generally much lower than their performance for less subtle lesions. The goal of this project is to develop a CAD system using advanced computer vision techniques aiming at improved detection of retrospectively seen cancers on prior mammograms and incorporate the developed CAD system into our current CAD system. During the project years, we have performed the following tasks: (1) collect the data sets of digitized film mammograms for training and testing our CAD system, (2) develop a series of single-view computer vision techniques for mass detection and classification in prior mammograms, (3) reduce FPs by correlation of image information from multiple view mammograms of the same patient, (4) develop a information fusion scheme to combine the new CAD system with the existing CAD system for mass detection, and (5) evaluate the effects of the newly developed CAD scheme with a large data set. We have found that our new computer-vision techniques can significantly improve the performance of the CAD system for mass detection by JAFROC analysis. The significance of this project is that the newly developed CAD system may be able to aid radiologists in detecting breast cancers at an early stage. Since early detection and treatment can reduce breast cancer mortality rate and health care costs, the proposed CAD system will improve the efficacy of mammography for breast cancer screening.

15. SUBJECT TERMS
Breast cancer, computer-aided detection, mammography

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU		USAMRMC

Table of Contents

(4) Introduction

Mammographic screening has been recognized as the most effective method for early detection of breast cancer[1-4]. Studies indicate that radiologists do not detect all carcinomas that are visible upon retrospective analyses of the images[5-10]. Various methods are being developed to improve the sensitivity and specificity of breast cancer detection. Double reading can reduce the miss rate of radiographic reading. However, double-reading by radiologists is costly. Computer-aided diagnosis (CAD) is considered to be one of the promising approaches that may improve the efficacy of mammography[11, 12]. It has been shown that CAD can improve radiologists' detection accuracy significantly[13-16]. Our receiver operating characteristic (ROC) study[17, 18] and that by Jiang et al.[17, 18] also showed that computer classifiers can improve radiologists' ability in differentiating malignant and benign masses or microcalcifications. CAD is thus a viable cost-effective alternative to double reading by radiologists.

Most of the CAD systems developed so far are based on radiologists' markers on mammograms which were proved to be cancer with biopsy. Some researchers[19-23] have investigated the performance change of CAD systems when using prior mammograms (i.e., the mammograms in previous exams on which the cancer can be seen retrospectively but was called negative or probably benign at the time of the exam). The ability of a CAD system to detect these cancers is important because it signifies early detection of cancers that radiologists may overlook. On the other hand, when a CAD system is applied to a new mammogram in clinical practice, it has to detect breast lesions of all degrees of subtlety effectively. Our experiences indicate that it is difficult to train a single CAD system to provide optimal detection for all lesions over the entire spectrum of subtlety because the classifiers have to make compromises to accommodate cancers of a wide range of characteristics.

The goal of this proposed project is to develop a CAD system using advanced computer vision techniques aiming at improved detection of retrospectively seen cancers on prior mammograms and incorporate the developed CAD system into our current CAD system. We hypothesize that a dual CAD system, which combines a system trained with subtle lesions retrospectively seen on prior mammograms and a system trained with cancers detected on current mammograms, should increase the sensitivity of detecting cancers at the early stage without compromising its ability to detect less subtle cancers. To accomplish this goal, we have performed the following tasks: (1) collection of a large database of masses on digitized prior and current film mammograms (DFMs) for training and testing the CAD system, (2) development of single-view computer vision techniques for mass detection and classification in prior DFMs, (3) reduction of false positives (FPs) by correlation of image information from multiple-view mammograms, (4) development of a dual system scheme which combines the new CAD system with our current CAD system without an increase in overall FPs, and (5) evaluation of the effects of the developed CAD with a large data set in detecting both average and subtle cancers.

It is expected that we will develop a fully automated CAD system which can be used for detection of masses on DFMs. Although we do not plan to develop such a system for digital mammograms because there will not be enough prior digital mammograms with missed cancers available for the development, the general methodology developed in this study can be adapted to CAD systems for digital mammograms in the future. The significance of this project is that have accomplished the goal of developing a CAD system which can further improve radiologists' accuracy in detecting breast cancers at an early stage. Since early detection and treatment can reduce breast cancer mortality rate, the CAD system will be useful for increasing the effectiveness of mammographic screening.

(5) Body

This is the final report of this project. We have described in detail the results of our studies in the past annual progress reports. The investigations conducted in this project are summarized in the following.

(5.A) Collection of a data set of mammograms (Task I)

With IRB approval, we have collected a database of digitized screen-film mammogram (DFM) from patient files in the Department of Radiology at the University of Michigan. In this study, we collected the mass data set contained 220 cases with masses. Each case included the current mammograms on which the mass was detected by radiologists, and the prior mammograms obtained from previous exams. The mass set contained 440 current mammograms and 496 prior mammograms. The true location of each mass was identified by an experienced Mammography Quality Standards Act (MQSA) radiologist. The radiologist also measured the mass size and provided descriptions of the mass margin, shape, conspicuity, and breast density.

(5.B) Development of single-view computer vision techniques for mass detection and classification on prior mammograms (Task II)

In this project, we have developed a series of computer vision techniques for mass detection on single-view mammograms. We have newly developed a two-stage gradient field analysis method which uses not only the shape information of masses on mammograms but also incorporates a second stage in which the gray level information of the local object segmented by a region growing technique is refined by gradient field analysis. In comparing with spatial gray level dependence (SGLD) texture features extracted from current mammograms, we extracted gray level features and run length statistics analysis (RLS) texture features inside and outside of the mass region on both the original image and gradient field image from prior mammograms. In CAD applications, an important step is to design a classifier for the differentiation of the abnormal from the normal structures. In this project, we have also investigated the performance of a regularized discriminant analysis (RDA) classifier in combination with a feature selection method for classification of the masses and normal tissues detected on mammograms. We have applied these computer vision techniques to mass detection on both full field digital mammograms (FFDM) and DFMs. We found that they were very useful and consistent for improving the accuracy of mass detection on both FFDMs and DFMs. (Publications: J1, J3, P1, P4, P5)

(5.C) Reduction of FPs by correlation of image information from multiple view mammograms (Task III)

In mammographic screening, a craniocaudal (CC) and a mediolateral oblique (MLO) or lateral (LAT) view are generally taken for each breast. The two views not only allow most of the breast tissue to be imaged but also improve the chance that a lesion will be seen in at least one of the views. The radiologist uses the two views to confirm true positives (TPs) and to reduce false positives (FPs). In an effort to improve the performance of our single CAD system, we have conducted several studies by using multiple view information of the same patient. We first developed a two-view information fusion method which combines the information from two mammographic views of the same breast. Then, we investigated an FP reduction method based on analysis of bilateral mammograms. We have also developed a four-view CAD system to improve the performance of mass detection for our computerized detection system. We found that our multiple view CAD system significantly improved the accuracy for mass detection on mammograms. (Publications J4, J5, P3, A3, A4, A7)

(5.D) Development of a information fusion scheme to combine the new CAD system with the existing CAD system for mass detection (Task IV)

In this project, we have developed a dual system scheme which combined a CAD system optimized with "average" masses with another CAD system optimized with "subtle" masses. The two single CAD systems for mass detection have similar image processing steps and were trained: one with the current mammograms and the other with the prior mammograms. A feed-forward backpropagation artificial neural network (BP-ANN) was trained to classify the masses from normal tissues by combining the output information from the two single CAD systems. In this ANN, the nodes are organized in an input layer, an output layer, and one hidden layers. The two linear discriminant analysis (LDA) scores from the two CAD systems were used as input to the BP-ANN. The BP-ANN has two input nodes, a single hidden layer with 3 hidden nodes, and one output node. The nodes are interconnected by weights and information propagates from one layer to the next through a log-sigmoidal transform function. The learning of the ANN is a supervised process in which known training cases are input to the ANN. The performance function for the feedforward network was the mean-square error which was the average squared error between the network outputs and the target values over all training samples. The gradient of the performance function was used to determine how to adjust the weights to minimize the error. The gradient is determined using an iterative backpropagation procedure which involves performing computations backward through the network. We found that the ANN fusion scheme can provide significant improvement in the accuracy of the mass detection CAD system in comparison with that of a single CAD system. (Publications J2, P2, A1, A2, A5, A6)

(5.E) Evaluation of the proposed CAD system with a large data set (Task V)

The detection of masses on mammograms is a challenging task because the overlapping fibroglandular tissue may mimic a mass or obscures the lesion. Although researchers have devoted extensive efforts to the development of CAD systems for mass detection, the performances of current CAD systems are far from ideal. We have developed a dual system approach and a four-view analysis method. In the end of this project, we have combined the dual system approach with the four-view approach and collected a relatively large data set to evaluate the effectiveness of our four-view dual CAD system. Besides the data set collected in this project, we also included 369 patients collected by our previous projects. In total, we used 589 patients including a mass set with 389 patients and a normal set with 200 patients in this study. Each patient had two views (CC and MLO/LAT) for each breast. The overall test performance was assessed by the free response receiver operating characteristic (FROC) curves. It was found that our the four-view dual CAD system achieved an FP rate of 1.04, 0.80, and 0.60 FPs/image at the case-based sensitivities of 90%, 85% and 80%, respectively, which represents a statistically significant improvement over the conventional single-view detection approach, using the Jackknife alternative FROC (JAFROC) method.. (Publications J5, A7)

(6) Key Research Accomplishments

- Collected 220 cases DFMs with 936 mammograms for training and testing the computerized detection systems (Task I).
- Developed a series of computer vision techniques for mass detection and classification on single-view mammograms (Task II).
- Developed a four-view CAD system for mass detection (Task III).
- Developed a dual system approach for mass detection (Task IV).

- Combined the dual system approach with four-view approach and evaluated the combined system with a relatively large data set (Task V).

(7) Reportable Outcomes

As a result of the support by the USAMRMC BCRP grant, we have developed a dual system approach with four-view analysis which significantly improved the performance of mass detection on mammograms in comparison with the conventional single-view detection approach. The publications from this project are listed below. Many of these have been reported in the previous annual reports.

Journal Articles:

J1. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, H. P., Petrick, N., Helvie, M. A., Roubidoux, M. A., Ge, J. and Zhou, C., "Computer aided detection of breast masses on full field digital mammograms," Medical Physics 32, 2827-2838 (2005).

J2. Wei, J., Chan, H.-P., Sahiner, B., Hadjiiski, L. M., Helvie, M. A., Roubidoux, M. A., Zhou, C. and Ge, J., "Dual system approach to computer-aided detection of breast masses on mammograms," Medical Physics 33, 4157-4168 (2006).

J3. Wei, J., Hadjiiski, L. M., Sahiner, B., Chan, H. P., Ge, J., Roubidoux, M. A., Helvie, M. A., Zhou, C., Wu, Y. T., Paramagul, C. and Zhang, Y., "Computer aided detection systems for breast masses: Comparison of performances on full-field digital mammograms and digitized screen-film mammograms," Academic Radiology 6, 659-669 (2007).

J4. Wu, Y.-T., Wei, J., Hadjiiski, L. M., Sahiner, B., Zhou, C., Ge, J., Shi, J., Zhang, Y. and Chan, H. P., "Bilateral analysis based false positive reduction for computer-aided mass detection," Medical Physics 34, 3334-3344 (2007).

J5. Wei, J., Chan, H.-P., Sahiner, B., Zhou, C., Hadjiiski, L. M., Roubidoux, M. A. and Helvie, M. A., "Computer-aided detection of breast masses on mammograms: Dual system approach with two-view analysis," Medical Physics (Accepted).

Conference Proceeding:

P1. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, H.-P., Petrick, N., Helvie, M. A., Zhou, C. and Ge, Z., "Computer aided detection of breast masses on full-field digital mammograms: False positive reduction using gradient field analysis," Proc. SPIE 5370, 992-998 (2004).

P2. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, H. P., A, H. M., A, R. M., Petrick, N., Zhou, C. and Ge, J., "Computer aided detection of breast masses on mammograms: Performance improvement using a dual system," Proc. SPIE 5747, 9-15 (2005).

P3. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, H.-P., Helvie, M. A., Roubidoux, M. A., Zhou, C., Ge, J. and Zhang, Y., "Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms," Proc. SPIE 6144, 241-247 (2006).

P4. Wei, J., Sahiner, B., Zhang, Y., Chan, H.-P., Hadjiiski, L. M., Zhou, C., Ge, J. and Wu, Y.-T., "Regularized discriminate analysis for breast mass detection on full field digital mammograms," SPIE Proc. 6144, 5P1~5P6 (2006).

P5. Wei, J., Sahiner, B., Chan, H. P., Hadjiiski, L. M., Roubidoux, M. A., Helvie, M. A., Ge, J., Zhou, C. and Wu, Y.-T., "Computer-aided detection of breast masses on prior mammograms," Proc. SPIE 6514, 651405-651407 (2007).

Conference Abstracts and Presentations:

A1. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, HP., Helvie, M. A., Roubidoux, M. A., "A dual computer-aided detection (CAD) system for improvement of mass detection on mammograms", Presentation at the 90th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL, November 28-December 3, 2004.

A2. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, HP., Helvie, M. A., Roubidoux, M. A., "A dual system for improvement of computer-aided mass detection on mammograms", Presentation at U. S. Army Medical Research and Materiel Command, Department of Defense, Breast Cancer Research Program, Phil., PA, June 8-11, 2005. Proceedings book p. 111.

A3. Wu, YT., Wei, J., Hadjiiski, L. M., Sahiner, B., Zhou , C., Chan, HP., "Bilateral CAD system for breast mass detection on mammograms", Presentation at the 92nd Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL. November 26-December 1, 2006.

A4. Wei, J., Sahiner, B., Chan, HP., Roubidoux, M. A., Helvie, M. A., Wu, YT., Hadjiiski, L. M., Ge, J. , Zhou, C., " Computer-aided detection of breast masses on mammograms: Performance improvement using two-view information", Presentation at the 92nd Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL. November 26-December 1, 2006.

A5. Wei, J., Chan, HP., Sahiner, B., Hadjiiski, L. M., Helvie, M. A., Roubidoux, M. A., Zhou, C., Ge, J., Wu, YT., Zhang, Y., "Dual system approach for improvement of subtle mass detection on mammograms: Improving prescreening with Hessian analysis", 93rd Scientific Assembly and Annual Meeting of the Radiological Society of North America, RSNA 2007.

A6. Wei, J., Chan, HP., Sahiner, B., Hadjiiski, L. M., Helvie, M. A., Roubidoux, M. A., "Dual-system two- view approach for computer-aided detection of breast masses on mammograms", Era of Hope 2008, Proceedings book p. 54.

A7. Wei, J., Chan, HP., Wu, YT., Sahiner, B., Hadjiiski, L. M., Helvie, M. A., C, Zhou, et.al., "Four-view Computer-aided Detection (CAD) System for Breast Masses on Mammograms", 94th Scientific Assembly and Annual Meeting of the Radiological Society of North America, RSNA 2008.

(8) Conclusions

In this project, we have developed a dual CAD system, which combined a CAD system optimized with "average" masses with another CAD system optimized with "subtle" masses, for mass detection on mammograms. Our studies showed that the improvement in the FROC curves

by the dual system approach was statistically significant ($p<0.05$) for the detection of both average masses and subtle masses using JAFROC method. In addition, we have developed a four-view approach to improve computerized detection of breast masses on mammograms. We have evaluated our approach by using a relatively large data set. Our results indicate that the proposed approach is able to further improve the detection performance as estimated by the JAFROC analysis. The significance of this project is that the newly developed CAD system may be able to aid radiologists in detecting breast cancers at an early stage. Since early detection and treatment can reduce breast cancer mortality rate and health care costs, the proposed CAD system will improve the efficacy of mammography for breast cancer screening.

(9) Personnel

Jun Wei, Ph.D.	PI
Berkman Sahiner, Ph.D.	Mentor
Mark A. Helvie, M.D.	Co-investigator
Marilyn A. Roubidoux, M.D.	Co-investigator
Heang-Ping Chan, Ph.D.	Consultant
Lubomir M. Hadjiiski, Ph.D.	Consultant

(10) References

[1] Smart, C. R., Hendrick, R. E., Rutledge, J. H. and Smith, R. A., "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," Cancer 75, 1619-1626 (1995).

[2] Byrne, C., Smart, C. R., Cherk, C. and Hartmann, W. H., "Survival advantage differences by age: Evaluation of the extended follow-up of the breast cancer detection demonstration project," Cancer 74, 301-310 (1994).

[3] Feig, S. A. and Hendrick, R. E., "Risk, benefit, and controversies in mammographic screening," *Syllabus: A categorical course in physics technical aspects of breast imaging*, A. G. Haus and M. J. Yaffe, Eds., 119-135, Radiological Society of North America, Inc (1993).

[4] Seidman, H., Gelb, S. K., Silverberg, E., Laverda, N. and Lubera, J. A., "Survival experience in the breast cancer detection demonstration project," CA Cancer J Clin. 37, 258-290 (1987).

[5] Hillman, B. J., Fajardo, L. L., Hunter, T. B., Mockbee, B., Cook, C. E., Hagaman, R. M., Bjelland, J. C., Frey, C. S. and Harris, C. J., "Mammogram interpretation by physician assistants," AJR 149, 907-911 (1987).

[6] Bassett, L. W., Bunnell, D. H., Jahanshahi, R., Gold, R. H., Arndt, R. D. and Linsman, J., "Breast cancer detection: One versus two views," Radiology 165, 95-97 (1987).

[7] Wallis, M. G., Walsh, M. T. and Lee, J. R., "A review of false negative mammography in a symptomatic population," Clinical Radiology 44, 13-15 (1991).

[8] Harvey, J. A., Fajardo, L. L. and Innis, C. A., "Previous mammograms in patients with impalpable breast carcinomas: Retrospective vs blinded interpretation," AJR 161, 1167-1172 (1993).
[9] Bird, R. E., Wallace, T. W. and Yankaskas, B. C., "Analysis of cancers missed at screening mammography," Radiology 184, 613-617 (1992).
[10] Beam, V., Sullivan, D. and Layde, P., "Effect of human variability on independent double reading in screening mammography," Academic Radiology 3, 891-897 (1996).
[11] Shtern, F., Stelling, C., Goldberg, B. and Hawkins, R., "Novel technologies in breast imaging: National cancer institute perspective," *2nd Postgraduate Course Syllabus*. 153-156, (1995).
[12] Vyborny, C. J., "Can computers help radiologists read mammograms?," Radiology 191, 315-317 (1994).
[13] Chan, H. P., Doi, K., Vyborny, C. J., Schmidt, R. A., Metz, C. E., Lam, K. L., Ogura, T., Wu, Y. and Macmahon, H., "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," Investigative Radiology 25, 1102-1110 (1990).
[14] Kegelmeyer, W. P., Pruneda, J. M., Bourland, P. D., Hillis, A., Riggs, M. W. and Nipper, M. L., "Computer-aided mammographic screening for spiculated lesions," Radiology 191, 331-337 (1994).
[15] Warren Burhenne, L. J., Wood, S. A., D'orsi, C. J., Feig, S. A., Kopans, D. B., O'shaughnessy, K. F., Sickles, E. A., Tabar, L., Vyborny, C. J. and Castellino, R. A., "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology 215, 554-562 (2000).
[16] Freer, T. W. and Ulissey, M. J., "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology 220, 781-786 (2001).
[17] Chan, H. P., Sahiner, B., Helvie, M. A., Petrick, N., Roubidoux, M. A., Wilson, T. E., Adler, D. D., Paramagul, C., Newman, J. S. and Gopal, S. S., "Improvement of radiologists' characterization of mammographic masses by computer-aided diagnosis: An roc study," Radiology 212, 817-827 (1999).
[18] Jiang, Y., Nishikawa, R. M., Schmidt, R. A., Metz, C. E., Giger, M. L. and Doi, K., "Improving breast cancer diagnosis with computer-aided diagnosis," Academic Radiology 6, 22-33 (1999).
[19] Nishikawa, R. M., Giger, M. L., Wolverton, D. E., Schmidt, R. A. and Kunio, D., "Prospective testing of a clinical cad workstation for the detection of breast lesions on mammograms," K. Doi, H. Macmahon, M. L. Giger, et al., Eds. Computer-Aided Diagnosis in Medical Imaging, 209-214, Elsevier Science, Amsterdam (1999).
[20] Petrick, N., Chan, H. P., Sahiner, B., Helvie, M. A., Paquerault, S. and Hadjiiski, L. M., "Breast cancer detection: Evaluation of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients.," Radiology 224, 217-224 (2002).
[21] Ikeda, D. M., Birdwell, R. L., O'shaughnessy, K. F. and Sickles, E. A., "Cad results on detectable but non-recalled findings for 169 prior "Negative" Mammograms in patients with subsequent screen-detected breast cancer," Radiology 213(P), 324 (1999).
[22] Birdwell, R. L., Ikeda, D. M., O'shaughnessy, K. F. and Sickles, E. A., "Evaluation of the potential of computer-aided detection (cad) to mitigate detection errors in screening mammography," Radiology 213(P), 150 (1999).
[23] Zheng, B., Good, W. F., Armfield, D. R., Cohen, C., Hertzberg, T., Sumkin, J. H. and Gur, D., "Performance change of mammographic cad schemes optimized with most-recent and prior image databases," Academic Radiology 10, 283-288 (2003).

(11) Appendix

The following publications are enclosed with this report:

Journal Articles:

J1. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, H. P., Petrick, N., Helvie, M. A., Roubidoux, M. A., Ge, J. and Zhou, C., "Computer aided detection of breast masses on full field digital mammograms," Medical Physics 32, 2827-2838 (2005).

J2. Wei, J., Chan, H.-P., Sahiner, B., Hadjiiski, L. M., Helvie, M. A., Roubidoux, M. A., Zhou, C. and Ge, J., "Dual system approach to computer-aided detection of breast masses on mammograms," Medical Physics 33, 4157-4168 (2006).

J3. Wei, J., Hadjiiski, L. M., Sahiner, B., Chan, H. P., Ge, J., Roubidoux, M. A., Helvie, M. A., Zhou, C., Wu, Y. T., Paramagul, C. and Zhang, Y., "Computer aided detection systems for breast masses: Comparison of performances on full-field digital mammograms and digitized screen-film mammograms," Academic Radiology 6, 659-669 (2007).

J4. Wu, Y.-T., Wei, J., Hadjiiski, L. M., Sahiner, B., Zhou, C., Ge, J., Shi, J., Zhang, Y. and Chan, H. P., "Bilateral analysis based false positive reduction for computer-aided mass detection," Medical Physics 34, 3334-3344 (2007).

J5. Wei, J., Chan, H.-P., Sahiner, B., Zhou, C., Hadjiiski, L. M., Roubidoux, M. A. and Helvie, M. A., "Computer-aided detection of breast masses on mammograms: Dual system approach with two-view analysis," Medical Physics (Accepted).

Conference Proceeding:

P1. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, H.-P., Petrick, N., Helvie, M. A., Zhou, C. and Ge, Z., "Computer aided detection of breast masses on full-field digital mammograms: False positive reduction using gradient field analysis," Proc. SPIE 5370, 992-998 (2004).

P2. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, H. P., A, H. M., A, R. M., Petrick, N., Zhou, C. and Ge, J., "Computer aided detection of breast masses on mammograms: Performance improvement using a dual system," Proc. SPIE 5747, 9-15 (2005).

P3. Wei, J., Sahiner, B., Hadjiiski, L. M., Chan, H.-P., Helvie, M. A., Roubidoux, M. A., Zhou, C., Ge, J. and Zhang, Y., "Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms," Proc. SPIE 6144, 241-247 (2006).

P4. Wei, J., Sahiner, B., Zhang, Y., Chan, H.-P., Hadjiiski, L. M., Zhou, C., Ge, J. and Wu, Y.-T., "Regularized discriminate analysis for breast mass detection on full field digital mammograms," SPIE Proc. 6144, 5P1~5P6 (2006).

P5. Wei, J., Sahiner, B., Chan, H. P., Hadjiiski, L. M., Roubidoux, M. A., Helvie, M. A., Ge, J., Zhou, C. and Wu, Y.-T., "Computer-aided detection of breast masses on prior mammograms," Proc. SPIE 6514, 651405-651407 (2007).

Computer-aided detection of breast masses on full field digital mammograms

Jun Wei,[a] Berkman Sahiner, Lubomir M. Hadjiiski, Heang-Ping Chan, Nicholas Petrick, Mark A. Helvie, Marilyn A. Roubidoux, Jun Ge, and Chuan Zhou
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109

(Received 7 February 2005; revised 16 June 2005; accepted for publication 16 June 2005; published 23 August 2005)

We are developing a computer-aided detection (CAD) system for breast masses on full field digital mammographic (FFDM) images. To develop a CAD system that is independent of the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM image as input and developed a multiresolution preprocessing scheme for image enhancement. A two-stage prescreening method that combines gradient field analysis with gray level information was developed to identify mass candidates on the processed images. The suspicious structure in each identified region was extracted by clustering-based region growing. Morphological and spatial gray-level dependence texture features were extracted for each suspicious object. Stepwise linear discriminant analysis (LDA) with simplex optimization was used to select the most useful features. Finally, rule-based and LDA classifiers were designed to differentiate masses from normal tissues. Two data sets were collected: a mass data set containing 110 cases of two-view mammograms with a total of 220 images, and a no-mass data set containing 90 cases of two-view mammograms with a total of 180 images. All cases were acquired with a GE Senographe 2000D FFDM system. The true locations of the masses were identified by an experienced radiologist. Free-response receiver operating characteristic analysis was used to evaluate the performance of the CAD system. It was found that our CAD system achieved a case-based sensitivity of 70%, 80%, and 90% at 0.72, 1.08, and 1.82 false positive (FP) marks/image on the mass data set. The FP rates on the no-mass data set were 0.85, 1.31, and 2.14 FP marks/image, respectively, at the corresponding sensitivities. This study demonstrated the usefulness of our CAD techniques for automated detection of masses on FFDM images. © *2005 American Association of Physicists in Medicine*.
[DOI: 10.1118/1.1997327]

Key words: computer-aided detection, full field digital mammogram (FFDM), multiresolution image enhancement, gradient field analysis, stepwise linear discriminant analysis

I. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 and 55 years of age.[1] It has been reported that early diagnosis and treatment can significantly improve the chance of survival for patients with breast cancer.[2–4] Although mammography is the best available screening tool for detection of breast cancers, studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially.[5–8] Computer-aided diagnosis (CAD) is considered to be one of the promising approaches that may improve the sensitivity of mammography.[9,10] Computer-aided lesion detection can be used during screening to reduce oversight of suspicious lesions that warrant further work-up. Computer-aided lesion characterization can assist in the estimation of the likelihood of malignancy of lesions by using image and/or other information during the diagnostic stage. The majority of studies to date show that CAD can improve radiologists' lesion detection sensitivity,[11–16] although Gur *et al.*[17] found that CAD had no significant effect on the radiologists in their academic setting when they averaged the results from both low-volume and high-volume radiologists. Further analysis of Gur's data by Feig *et al.*[18] indicated that the 17 low-volume radiologists in Gur's study achieved similar increase in sensitivity as reported in other studies. The outcome of CAD studies therefore depends on the study design and data analysis.

A number of investigators have reported CAD algorithms for detection of masses on mammograms. Their approaches to prescreening of mass candidates were based primarily on mass characteristics including: (1) asymmetric density between left and right mammograms,[19–22] (2) texture,[23,24] (3) spiculation,[25,26] (4) gray level contrast,[27–31] and (5) gradient.[32] Some of these approaches were refined with a combination of the mass characteristics. Feature classifiers were then used to further differentiate masses from normal breast tissues.

Most mammographic CAD algorithms developed so far are based on digitized screen-film mammograms (SFMs). In the last few years, full field digital mammographic (FFDM) technology has advanced rapidly because of the potential of digital imaging to improve breast cancer detection. Several manufacturers have obtained clearance from the FDA for clinical use. It is expected that FFDM detectors will provide higher signal-to-noise ratio (SNR) and detective quantum efficiency, wider dynamic range, and higher contrast sensitivity

than digitized mammograms. The spatial resolution of digital detectors may also be different from that of digitized SFMs even when their pixel pitches are equal. Li *et al.* investigated the performance of their CAD system on mass detection that was developed for SFMs and modified for FFDMs.[33] Their preliminary results on a small data set showed that it achieved 60% sensitivity at 2.47 false positives (FPs)/image. It is expected that proper adaptation based on the imaging characteristics of FFDMs and re-training of the CAD system with FFDMs would improve the performance. Because of the higher SNR and linear response of digital detectors, there is also a strong potential that more effective feature extraction techniques can be designed to optimally extract signals from the image and improve the accuracy of CAD. Several commercial CAD systems already obtained FDA approval for use with FFDMs. The commercial CAD systems generally reported similar performance on FFDMs and SFMs. However, their study was not reported in peer-reviewed journals so that the data set and algorithm are unknown. Recently, an assessment study[34] to compare the performance of two commercial and one research CAD system for SFMs showed that their mass detection sensitivities ranged from 67% to 72% and the FP rates ranged from 1.08 to 1.68 per four-view examinations. The differences in sensitivities were not significant whereas the differences in the FP rates were significant, depending on the examinations and CAD systems used.[34]

We have developed a CAD system for the detection of masses on SFMs in our previous studies.[30,35,36] We are developing a mass detection system for mammograms acquired directly by a FFDM system. In this study, we adapted our mass detection system developed for SFMs to FFDMs by optimizing each stage and retraining. In an effort to develop a CAD system that is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM as input and developed a multiresolution preprocessing scheme for image enhancement. A new technique was also designed for prescreening of mass candidates on the preprocessed images.

II. MATERIALS AND METHOD

A. Data sets

The mammograms were collected from patient files at the Department of Radiology with Institutional Review Board approval. Digital mammograms at the University of Michigan are acquired with a GE Senographe 2000D FFDM system. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of 100 μm $\times$100 μm and 14 bits per pixel. In this study, we used two data sets: a mass set containing FFDMs with malignant or benign masses and a no-mass set containing FFDMs without masses. The no-mass set was obtained from microcalcification cases collected for the development of our microcalcification CAD systems. The cases were included as normal, with respect to masses, only if they were verified to be free of masses by an experienced Mammography Quality Standards Act (MQSA) radiologist. Our mass detection system aims at application to screening mammography so that the mass cases, regardless of malignant or benign, are considered positive. All cases had two mammographic views, the craniocaudal view and the mediolateral oblique view or the lateral (LM or ML) view. The mass set contained 110 cases with a total of 220 images. The no-mass set contained 90 cases with a total of 180 images. The mass data set was used to estimate the detection sensitivity and the no-mass data set was used for estimating the FP rate. There were a total of 110 biopsy-proven masses in the mass data set. Eighty-seven of the masses were benign and 23 of the masses were malignant. A MQSA radiologist identified the locations of the masses, measured the mass sizes as the longest dimension seen on the two-view mammograms, provided descriptors of the mass shapes and mass margins, and also provided an estimate of the breast density in terms of BI-RADS category. Figure 1 shows the information of our data set which includes the distributions of mass sizes, mass shapes, mass margins, and breast density.

B. Methods

Our CAD system consists of five processing steps: (1) preprocessing by using multiscale enhancement, (2) prescreening of mass candidates, (3) identification of suspicious objects, (4) feature extraction and analysis, and (5) FP reduction by classification of normal tissue structures and masses. The block diagram for the detection scheme is shown in Fig. 2. These steps are described in more detail in the following.

We randomly separated the mass data set into two independent, equal sized subsets. Each subset contained 55 cases with 110 images. Cross validation was used for training and testing the algorithms. The training included selecting the preprocessing Laplacian pyramid reconstruction weights, adjusting the filter weights for prescreening and clustering, determining thresholds for rule-based classification, and selecting morphological and texture features and classifier weights. Once the training with one subset was completed, the parameters and all thresholds were fixed for testing with the other subset. The training and test subsets were switched and the training process was repeated. The overall detection performance was evaluated by combining the performances for the two test subsets. The trained algorithms with the fixed parameters were also applied to the no-mass mammograms to estimate the FP rate in screening mammograms.

1. Preprocessing

FFDMs are generally preprocessed with proprietary methods by the manufacturer of the FFDM system before being displayed to readers. The image preprocessing method used depends on the manufacturer of the FFDM system. To develop a CAD system that is less dependent on the FFDM manufacturer″s proprietary preprocessing methods, we use the raw FFDM as input to our CAD system. We developed a multiscale preprocessing scheme for image enhancement.

Multiscale methods have been used for contrast enhancement of medical images. Since a multiscale method uses the information from a large number of frequency channels ex-

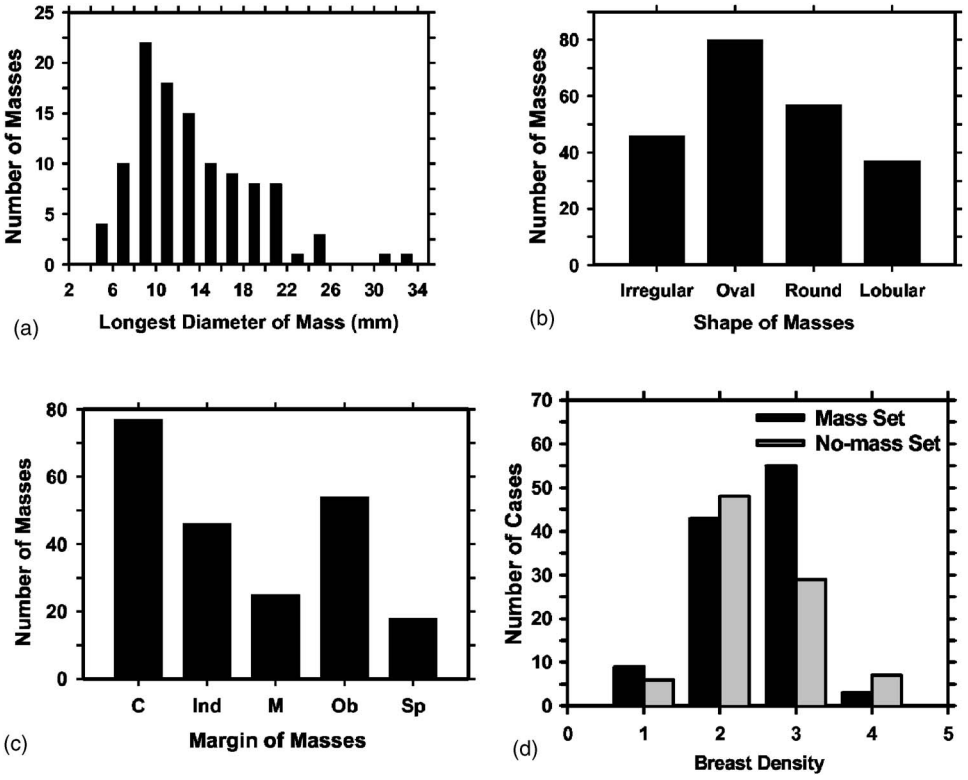


FIG. 1. The information of our mass data set: (a) distribution of mass sizes, (b) distribution of mass shapes, (c) distribution of mass margins, C: circumscribed, Ind: indistinct, M: microlobulated, Ob: obscured, Sp: spiqulated, (d) distribution of the breast density in terms of BI-RADS category estimated by a MQSA radiologist.

tracted from the image adaptively, it is more flexible and versatile than the commonly used enhancement methods, such as unsharp masking, which uses a small number of frequency channels. Two types of multiscale methods have been used as the preprocessing methods for the contrast enhancement of mammograms: the wavelet method and the Laplacian pyramid method.[37] A previous study has shown that, for the purpose of image enhancement, using a Laplacian pyramid method is advantageous compared to using the fast wavelet transformation which introduces visible artifacts.[38] In this project, therefore, we chose the Laplacian pyramid method as our preprocessing method.

A flowchart of our preprocessing method is shown in Fig. 3. In brief, the mammogram is first segmented automatically into the background and the breast region. Second, a logarithmic transform is applied to the breast image. The Laplacian pyramid method is used to decompose the breast image

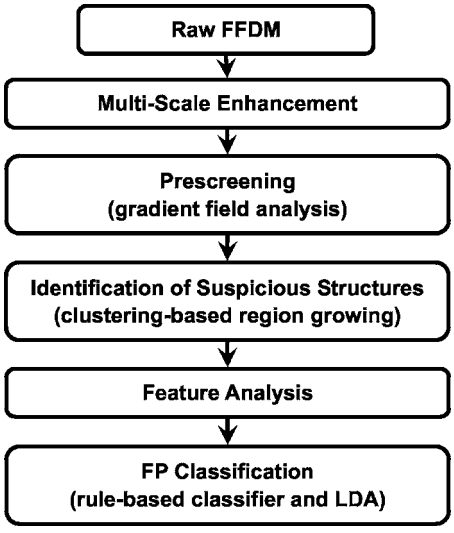


FIG. 2. Schematic diagram of our CAD system for mass detection on FFDM. The system is developed for screening mammography so that all masses, regardless of malignant or benign, are considered positive. The FP classification stage includes rule-based classification, a morphological LDA classifier, and a texture feature LDA classifier for differentiating masses from normal breast tissues.

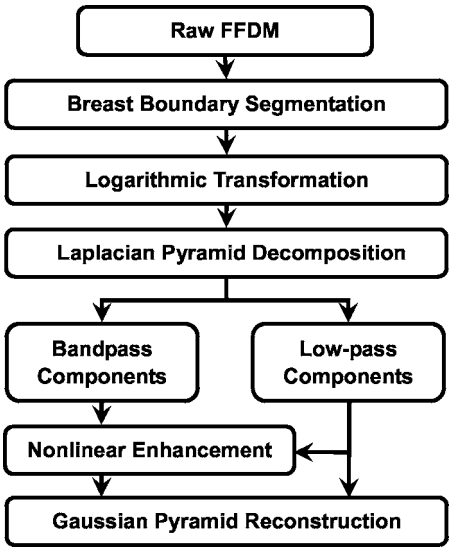


FIG. 3. Schematic diagram for the image preprocessing stage of our mass detection system, which includes breast boundary segmentation, logarithmic image transformation, and Laplacian pyramid multiscale enhancement.

into multiscales. A nonlinear weight function based on the pixel gray level from each of the low-pass components is designed to enhance the high-pass components.

Since the contrast between the breast and the background in a raw FFDM is high, a two-step algorithm was developed for the segmentation of breast region. First, Otsu's method[39] is used to calculate a threshold and binarize the original image. Second, an eight-connectivity labeling method is used to identify the connected regions below the threshold on the binary image. The region with the largest area will be considered to be the breast region.

Clinical mammograms are usually viewed in a negative mode of the raw images. In order to process an image with the same format as the clinical mammograms, we first use an inverted logarithmic function[40] to transform the raw data. A multiresolution method is then used to enhance the log-transformed image. The inverted logarithmic function for signal transfer can be expressed as

$$S_x = \ln\left(\frac{X_{\max}}{X}\right) \tag{1}$$

where X is the gray level of the raw data, $X_{\max}$ is the maximum of the 14 bit digital gray scale number (i.e., 16 383). The transformed image is then linearly scaled to 12 bit pixel values.

The Laplacian pyramid decomposition is a multiscale method that was first introduced as an image compression technique.[37] We previously evaluated the effect of Laplacian pyramid data compression on the detection of microcalcifications on digitized mammograms.[41] An illustration of a Laplacian decomposition tree is shown on the left-hand side of Fig. 4. The Laplacian pyramid is a sequence of error images $L_0, L_1, \ldots, L_n$. Each is the difference between two consecutive levels of the Gaussian pyramid $G_0, G_1, \ldots, G_n$, where G_0 is the original image. Each subsequent level of the Gaussian pyramid in the decomposition tree is generated by convolution of the image at the previous level with a 5×5 kernel, $w(m,n)$, that has weights of 0.4 at the center, 0.25 at the eight nearest neighbors of the center, and 0.05 at the 16 peripheral pixels, and then downsampled by a factor of 2, as described in Eq. (4). The decomposition of the image from level k to level $k+1$ can be expressed mathematically by

$$L_k = G_k - \text{Expand}(G_{k+1}), \tag{2}$$

where

$$\text{Expand}(G_{k+1}) = 4\sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n)\cdot G_{k+1}\left(\frac{i-m}{2},\frac{j-n}{2}\right), \tag{3}$$

$$G_k(i,j) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n)G_{k-1}(2i+m,2j+n). \tag{4}$$

The original image can be recovered by following the Gaussian reconstruction tree shown on the right-hand side of Fig. 4 if no enhancement is applied to the Laplacian pyramid. At a given level of the Gaussian reconstruction tree, the image is expanded (convolved and upsampled), as shown in Eq. (3), and then added to the Laplacian error image of the corresponding level. Details of the decomposition and reconstruction processes can be found in the literature.[37]

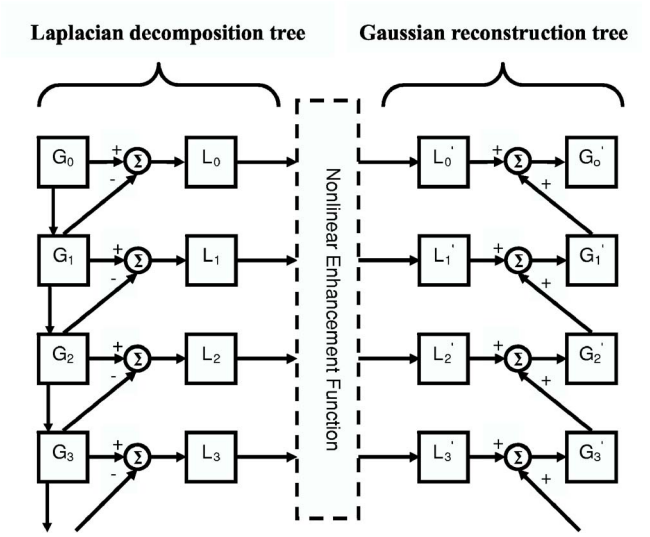


FIG. 4. Multiscale enhancement using the Laplacian pyramid decomposition method: Laplacian decomposition tree on the left-hand side and the Gaussian reconstruction tree on the right-hand side. The different levels of the Gaussian pyramid images are denoted by G_i, $(i=0,\ldots,n)$. The error images at different levels of the Laplacian pyramid are denoted by L_i, $(i=0,\ldots,n)$. The primed quantities G_i' and L_i' denoted the images at different levels after enhancement. Σ denotes the summation operation. The image is downsampled by a factor of 2 when it goes down every level of the decomposition tree, and upsampled by a factor 2 when it moves up every level of the reconstruction tree.

We enhance the reconstructed image to facilitate mass detection. The image at each level of the Laplacian pyramid that corresponds to a bandpass image is mapped by a nonlinear function. In this study, we use a nonlinear function that incorporates the information from each bandpass image. A Gaussian pyramid expansion is then used to reconstruct the image from the low pass components and the enhanced bandpass components, as shown in Fig. 4. The reconstruction scheme is defined by

$$r(k) = \alpha\cdot\text{Expand}(G_{k+1}) + \beta\cdot(\text{Expand}(G_{k+1}))^p\cdot L_k, \tag{5}$$

where α, β, and p are constant values in the range of 0.2–2.0 experimentally chosen for each frequency level.

Figures 5(a) and 5(b) show an example of a GE raw image and its processed image provided by the GE FFDM system. The histograms of the raw image and the processed image are shown next to the corresponding images. An example of the processed image using our multiresolution enhancement method and the corresponding histogram are shown in Fig. 5(c).

2. Prescreening and segmentation of suspicious objects

In our previous CAD system developed for digitized SFMs, an adaptive density-weighted contrast enhancement (DWCE) filter[35] was developed for prescreening. Although

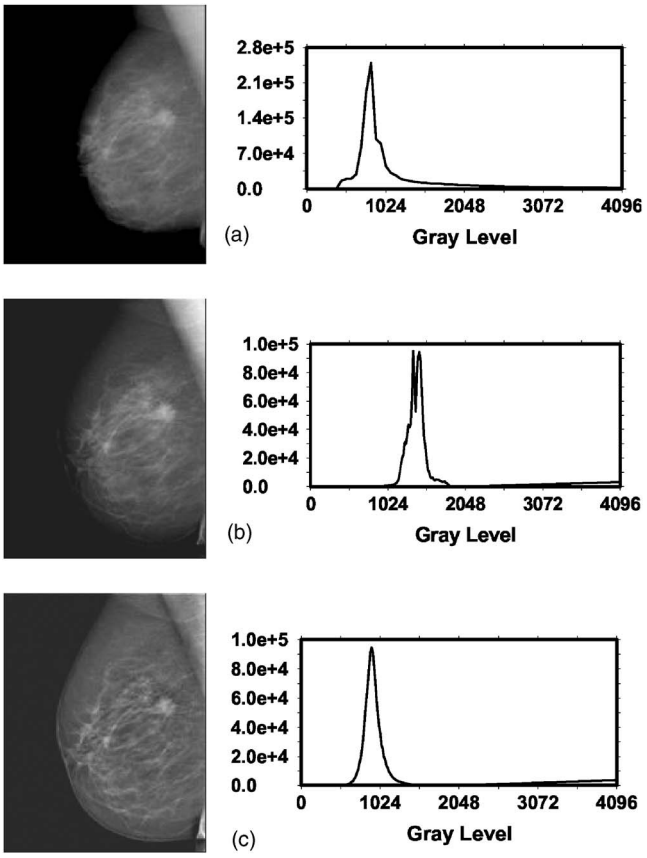


FIG. 5. An example of (a) GE raw image, (b) GE processed image, and (c) our processed image by using the Laplacian pyramid multiscale method. The gray level histogram of each image is also shown. The GE raw image has 14 bit gray levels but the histogram only plotted the lower 12 bits because very few pixels had gray levels higher than 4095.

the DWCE filter using the gray level information can identify the suspicious locations of masses on mammograms with high sensitivity, the prescreening objects often include a large number of enhanced normal breast structures.

In this study, we investigated the use of a new method that combines gradient field information and gray level information to detect mass candidates on FFDMs. Gradient field information is commonly used in computer vision or other fields to extract objects or intensity field distributions. Kobatake *et al.*[42] designed a filter, referred to as an iris filter, to calculate the convergence of gradient index around each pixel on SFMs which provided shape information for detection of masses. An extension of the iris filter, referred to as an adaptive ring filter, was developed by Wei *et al.*[43] for detection of lung nodules on chest x-ray images. In this study, we have developed a two-stage gradient field analysis method which uses not only the shape information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis.

To reduce noise in the gradient calculation, the image is smoothed with a 4×4 box filter and subsampled to 400 μm $\times$ 400 μm. The gradient field analysis is applied to the smoothed image. At each pixel $c(i)$ within the breast, concentric annular regions centered at $c(i)$ with an average radius, $R(k)$, of k pixels from $c(i)$ and a radial width of 4 pixels are defined within a circular region of about 12 mm in radius. The gradient vector at each pixel $p(j)$ within an annular region is computed and the gradient direction is obtained by projecting the gradient vector to the radial direction vector from $c(i)$ to $p(j)$. The average gradient direction over an annular region at the average radius $R(k)$ is calculated as the mean of the gradient directions over pixels on three adjacent annular regions $R(k-1)$, $R(k)$, and $R(k+1)$. Finally, the gradient field convergence at $c(i)$ was determined as the maximum of the average gradient directions among all annular regions. A region of interest (ROI) of 256×256 pixels in the 100 μm $\times$ 100 μm images is identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method[44] in which the location of high gradient convergence is used as the starting point. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. Objects whose new gradient convergence is lower than 80% of the original value are rejected.

After prescreening, the suspicious objects are identified by using a two-stage segmentation method. First, the background-corrected ROI was weighted by a Gaussian function with $\sigma=256$ pixels. Then, a k-means clustering using the pixel values in a background-corrected image and a Sobel filtered image as features is used to find the object. Figures 6(a) and 6(b) show the initial detection locations and the grown objects, respectively, obtained by prescreening the mammogram shown in Fig. 5(c).

3. Feature extraction and FP reduction

FP classification in our mass detection system is accomplished by a three-stage classification scheme.[36,44] For each suspicious object, eleven morphological features are extracted. Rule-based classification and a linear discriminant analysis (LDA) classifier using all 11 morphological features as input predictor variables are trained to remove the detected structures that are substantially different from breast masses. The training data set alone was used for training the classification rules and the weights of the LDA classifier. After morphological classification, global and local multiresolution texture analyses[45] are performed in each remaining ROI by using the spatial gray level dependence (SGLD) matrix. Briefly, the wavelet transform is employed to decompose an ROI into three levels for global texture analysis. Thirteen types of texture features[44,46] are extracted from each ROI. Each feature is calculated at 14 pixel distances and 2 angular directions. A total of 364 features (13 texture measures $\times$ 14 distances $\times$ 2 directions) is extracted from global texture analysis. Local texture features are extracted from the local region containing the detected object (object region) and the peripheral regions within each ROI. A total of 208 features (104 features from the object region and 104

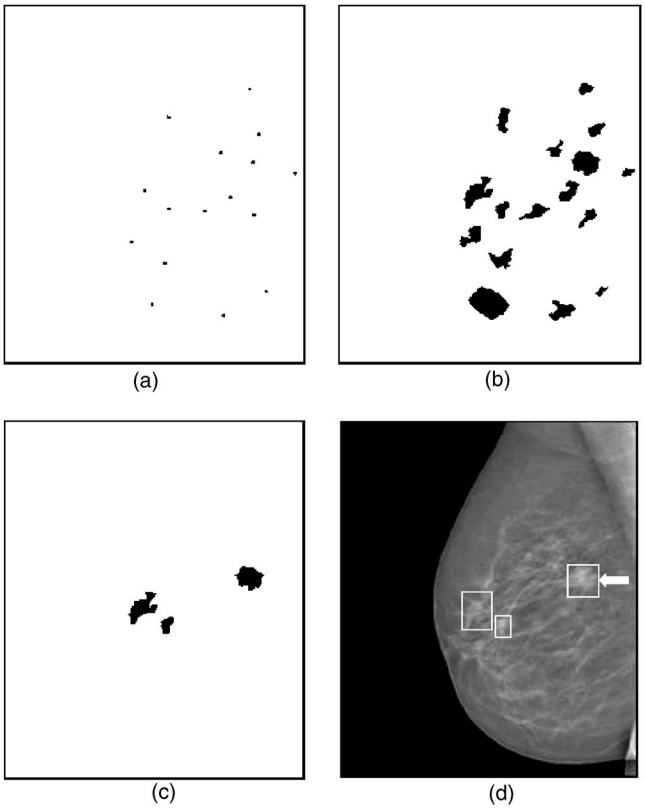


FIG. 6. An example demonstrating the processing steps with our CAD system: (a) object locations identified in prescreening, (b) identified suspicious objects, (c) detected objects after FP reduction, and (d) image superimposed with ROIs identifying the detected objects. The true mass is indicated by an arrow.

features from the peripheral regions) are extracted. The third-stage FP reduction using the texture features is described next.

4. Texture classification of masses and normal tissue

In order to obtain the best texture feature subset and reduce the dimensionality of the feature space to design an effective classifier, feature selection with stepwise LDA was applied. At each step one feature was entered or removed from the feature pool by analyzing its effect on the selection criterion, which was chosen to be the Wilks' lambda in this study. The optimization procedure used a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for excluding features that had high correlation with the features already in the selected pool. Since the appropriate values of F_{in}, F_{out}, and T were unknown, we examined a range of F_{in}, F_{out}, and T values using an automated simplex optimization method. For a given combination of F_{in}, F_{out}, and T values, the algorithm used a leave-one-case-out resampling method within the training subset to select features and estimate the weights for the LDA classifier. To evaluate the classifier performance, the test discriminant scores from the left-out cases were analyzed using receiver operating characteristic (ROC) methodology.[47] The discriminant scores of the mass and normal tissue were used as the decision variable in the LABROC program, which fits a binormal ROC curve based on maximum likelihood estimation. The accuracy for classification of mass and normal tissue was evaluated as the area under the ROC curve, A_z. The test A_z for the left-out cases in the leave-one-out resampling within the training subset was used as a figure of merit to guide the simplex algorithm to search for the best set of F_{in}, F_{out}, and T values within the parameter space. In this approach, feature selection was performed without the left-out case so that the test performance would be less optimistically biased.[48] However, the selected feature set in each leave-one-case-out cycle could be slightly different because every cycle had one training case different from the other cycles. In order to obtain a single trained classifier to apply to the test subset, a final stepwise feature selection was performed with the entire training subset and a set of F_{in}, F_{out}, and T thresholds chosen from the output of simplex training process. This set of F_{in}, F_{out}, and T thresholds was chosen based not only on the test A_z values, which were generated when the simplex procedure was searching through the parameter space, but also on the average number of features selected. The appropriate thresholds were chosen as a balance between keeping the number of selected features small and a relatively high classification accuracy by LDA. The chosen thresholds were then applied to the entire training subset to obtain the final set of features using stepwise feature selection and estimate the weights of the LDA. The LDA classifier with the selected feature set was then fixed and applied to the test subset. The test subset was independent of the training subset as described in Sec. II B 2 and was not used in the above-described leave-one-case-out classifier training process.

5. Evaluation methods

The detected individual objects were compared with the "truth" ROI marked by an experienced radiologist. A detected object was scored as true positive (TP) if the overlap between the bounding box of the detected object and the truth ROI was over 25%. Otherwise, it would be scored as FP. The 25% threshold was selected as described in our previous study.[36] The detection performance of the CAD system was assessed by free response ROC (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to 220 masses. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either or both masses on the two views was considered to be a TP detection; the sensitivity was thus calculated relative to 110 masses. Figure 6(c) shows an example of the final detected objects and Fig. 6(d) shows the locations of these objects superimposed on the mammogram.

To evaluate the effect of the preprocessing methods on mass detection, we also trained a CAD system using the GE

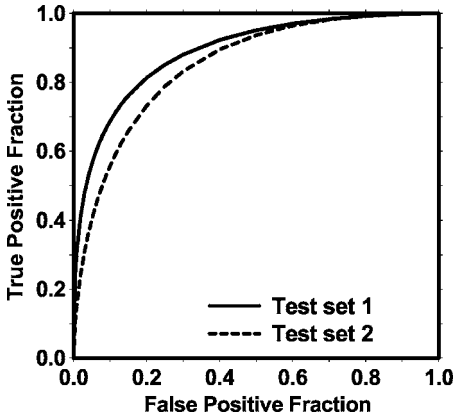


FIG. 7. The test ROC curves from the two independent mass subsets. The LDA classifiers using text features achieved an A_z value of 0.89 ± 0.02 for test subset 1 and 0.85 ± 0.02 for test subset 2 in the classification of mass and normal breast tissues.

processed image as input. This CAD system used the same methods as those described earlier for the raw images except that the Laplacian pyramid preprocessing step was not applied to the GE processed image, and that the prescreening and feature classifiers were retrained specifically for the GE processed images to obtain the best performance. The training and test subsets contained the same corresponding cases as for the raw image subsets. The training and testing were performed using the above-described cross validation method. The performance of the CAD system using the GE processed images was quantified by the average test FROC curve and compared with that using the raw images.

III. RESULTS

With raw images as input and Laplacian pyramid enhancement, our CAD system using the two-stage gradient field analysis detected 92.7% (204/220) of the masses with an average of 18.9 (4152/220) objects/image at the prescreening stage, compared with an average of 23.8 objects/image at the same sensitivity by using gradient field information alone. After FP reduction using the rule-based and linear classifier based on morphological features, there were a total of 3412 mass candidates (15.5 objects/image) at a sensitivity of 90.5% (199/220).

The texture-based LDA classifier for FP reduction was designed with stepwise feature selection and simplex optimization. The most effective subset of features from the available feature pool was selected for each of the training subsets during the training procedure. Twenty (11 global and 9 local) and 19 (12 global and 7 local) texture features were selected from the two independent training subsets, respectively. The test ROC curves are shown in Fig. 7. The training A_z values of the LDA classifier on the two training subsets were 0.87 ± 0.02 and 0.88 ± 0.01, respectively. The classifiers achieved A_z values of 0.89 ± 0.02 and 0.85 ± 0.02 on the independent test subsets, respectively. Figure 8 shows the FROC curves for the two test subsets after FP reduction with the corresponding trained LDA classifiers. An average FROC curve was derived from these two FROC curves by averaging the FP/images at the corresponding sensitivities. This average test FROC curve is plotted in Fig. 9 for comparison with the other FROC curves, described next.

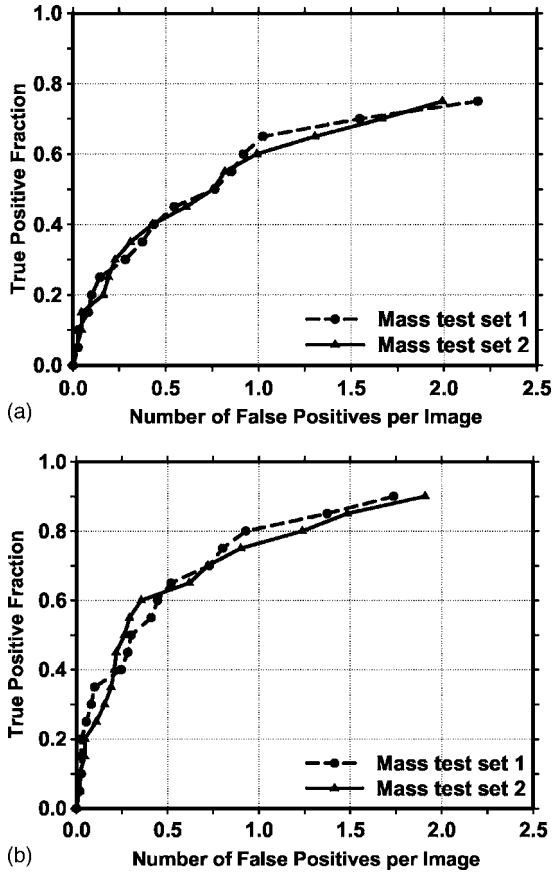


FIG. 8. The test FROC curves from the two independent mass subsets for the CAD system using the raw images as input and processed with the Laplacian pyramid method. The FP rate was estimated from the mammograms with masses. (a) Image-based FROC curves, (b) case-based FROC curves.

In addition to using the mass data set containing 110 cases for the cross validation training and testing, we used a no-mass data set containing 90 cases with 180 images to evaluate the FP detection rate in normal cases. Since two sets of trained parameters were acquired as a result of the cross validation training, we applied the two trained CAD systems separately to the no-mass data set for FP detection. The number of FP marks produced by the algorithm was determined by counting the detected objects on these normal cases only. The mass detection sensitivity was determined by counting only the abnormal objects on each of the test mass subsets. The combination of the sensitivity from each of the test mass subsets and the FP rate from the normal data set at the corresponding detection thresholds resulted in a test FROC curve. The two test FROC curves were then averaged, as described earlier, to obtain an overall FROC curve quantifying the test performance of the CAD system. Figures 9(a) and 9(b) show the comparison of the average FROC curves with the FP rates estimated from the two data sets. The test FROC curve with the FP rate estimated from the no-mass data set showed a case-based detection sensitivity of 70%,

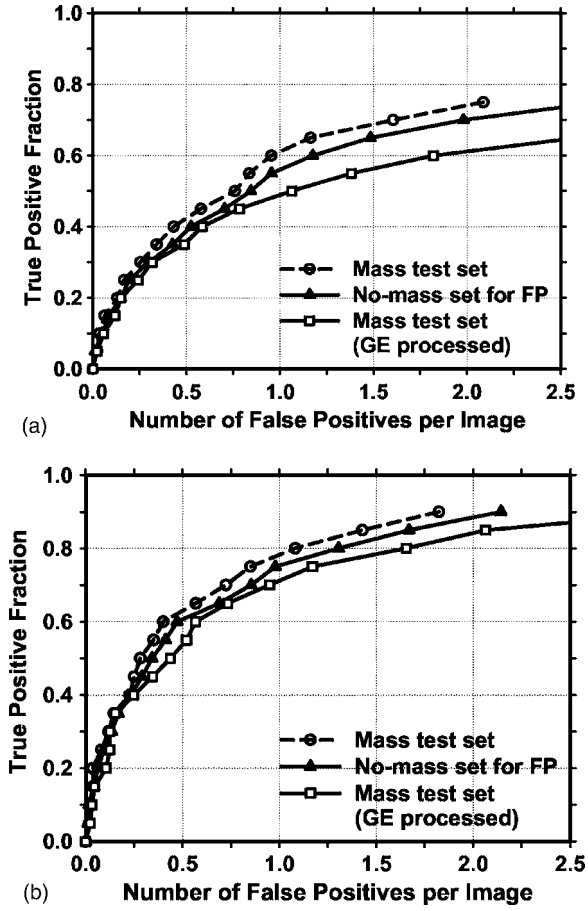


FIG. 9. Comparison of the average test FROC curves obtained from: (1) the CAD system using raw images as input, with the FP rate estimated from the mammograms with masses, (2) the CAD system using raw images as input, with the FP rate estimated from the normal mammograms without masses, and (3) the CAD system using GE processed images as input, with the FP rate estimated from the GE processed mammograms with masses. (a) Image-based FROC curves, (b) case-based FROC curves.

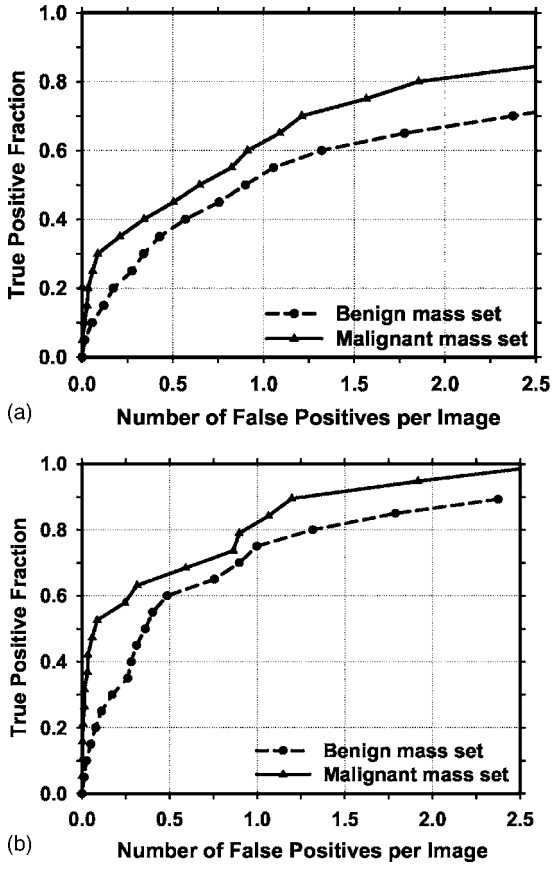


FIG. 10. Comparison of the average test FROC curves for the malignant and benign mass sets. The CAD system using raw images as input was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) case-based FROC curves.

80%, and 90% at 0.85, 1.31, and 2.14 FP marks/image, which are slightly higher than the FP rates of 0.7, 1.1, and 1.8 marks/image, respectively, estimated from the mass data set. Since our mass detection algorithm limits the maximum number of output marks to be 3 at the final stage, the FP marker rates will be slightly higher if the detection is performed in no-mass images. However, many images do not reach the maximum of 3 marks so that the difference in the FP marker rate between the mass and no-mass set is less than one. We also analyzed the detection accuracy of the system for malignant and benign masses separately. Figures 10(a) and 10(b) show the average FROC curves for detection of malignant and benign masses.

The average test FROC curves of the CAD system using the GE processed images as input were compared to those of the CAD system using raw images as input and Laplacian pyramid multiscale preprocessing as shown in Fig. 9. The FROC curves were plotted as the detection sensitivity as a function of the number of FP marks per image on the mass data set. The CAD system using the GE processed images as input achieved a case-based sensitivity of 70%, 80%, and 90% at 0.9, 1.6, and 3.1 FP marks/image, respectively, compared with 0.7, 1.1, and 1.8 FP marks/image on the CAD system using raw images as input.

IV. DISCUSSION

Several FFDM systems have been approved for clinical applications. It is important to develop a CAD system that can easily be adapted to images acquired by FFDM systems from different manufacturers. In this study, we are developing a CAD system that uses the raw FFDMs as the input. Since digital detectors generally have a linear response to x-ray exposure, the raw pixel values are a linear function of the absorbed x-ray energy in the detector. The signal range between different digital detectors can therefore be normalized linearly with respect to each other. Although the spatial resolution and noise properties of the images from different detectors are still different, the use of raw images already reduces one of the major differences between mammograms from different FFDM systems. For preprocessing of the raw images, we developed a multiresolution enhancement method. An example of a typical mammogram processed by the GE method and our method is compared in Fig. 5. As seen from this example, the enhancement of mammographic structures was stronger for our processed image than for the

TABLE I. Estimation of the statistical significance in the difference between the FROC performance of the CAD system using the FFDM raw images as input and processed with our Laplacian pyramid method and that of the CAD system using GE processed images as input. The FROC curves with the FP rates obtained from the no-mass data set (Fig. 9) were compared.

	A_1 (AFROC)			FOM (JAFROC)		
	Test subset 1	Test subset 2	p values	Test subset 1	Test subset 2	p values
Raw+LP processed	0.44	0.39	0.012	0.46	0.41	0.006
GE processed	0.37	0.31	0.0009	0.39	0.34	0.012

GE processed image. From a comparison of their histograms, it was found that the two histograms are very similar except for the average gray level.

For the evaluation of the effect of the preprocessing methods on computerized mass detection, we observed that our Laplacian pyramid preprocessing method provided higher detection accuracy than the GE processing method. As shown in Fig. 5, the Laplacian pyramid preprocessing method applies a stronger edge enhancement to the image than the GE method. Our preprocessing method aims at enhancing the image structures for computer vision whereas the GE processing method was designed to enhance the image for human visual interpretation. The stronger enhancement used for preprocessing the raw images appeared to improve the accuracy of the computer in detecting the masses.

Currently, there is no established statistical analysis method for testing the significance of the difference between two FROC curves generated by a CAD system. Chakraborty *et al.* proposed using an alternative free-response ROC (AFROC) method[49] to transform the FROC data to AFROC data, to which the curve fitting software and statistical significance tests for ROC analysis can then be applied and demonstrated its application to human observer performance rating data. In the AFROC method, false-positive images (FPIs) instead of FPs per image are counted. The confidence rating of a FPI is determined by the highest confidence FP decision on the image regardless of how many lower confidence FP decisions are made on the same image. We applied the AFROC method to evaluate the differences in pairs of our FROC curves that used the no-mass set for estimation of the FP rates. The ROCKIT software developed by Metz *et al.*[47] was used to analyze the AFROC data. The comparison of A_1 and p values is summarized in Table I. The area under the fitted AFROC curve (A_1) was 0.44 and 0.39, respectively, on mass test subsets 1 and 2 for the CAD system using raw images as input and processed with our Laplacian pyramid method, and 0.37 and 0.31, respectively, on the same subsets for the CAD system using GE processed images as input. The difference between the fitted AFROC curve for our processed images and that for the GE processed images was statistically significant ($p<0.05$) for both test subsets. However, all four fitted AFROC curves deviated systematically from the AFROC data (see two examples plotted in Fig. 11 for the test subset 1). It is uncertain whether the AFROC method is applicable to our FROC data and thus whether the statistical significance testing is valid.

More recently, Chakraborty *et al.*[50] proposed a JAFROC method and provided software to estimate the statistical significance of the difference between two FROC curves. We also applied the JAFROC analysis to the two pairs of FROC curves. The figure-of-merit (FOM) from the output of the JAFROC software was 0.46 and 0.41, respectively, on mass test subsets 1 and 2 for the CAD system using raw images as input and processed with our Laplacian pyramid method, and 0.39 and 0.34, respectively, on the same subsets for the CAD system using GE processed images as input. The difference between the FOM for our processed images and that for the GE processed images was again statistically significant ($p<0.05$). The FOM values were about 0.02 higher than the corresponding A_1 values. The JAFROC software did not provide a fitted curve or a goodness-of-fit indicator in the output so that it is not known whether this model fits our FROC data better than the AFRPC method. Although both methods indicate that the improvement in the FROC performance using our Laplacian pyramid processed images is statistically

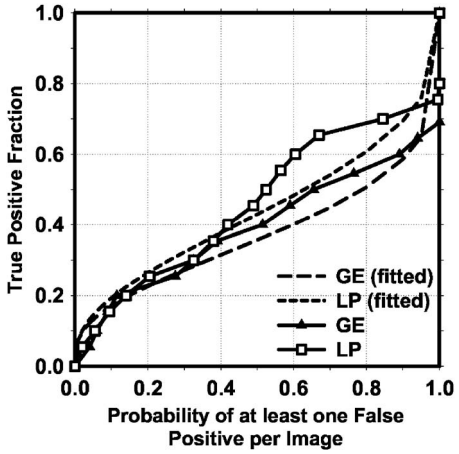


FIG. 11. Comparison of alternative free-response receiver operating characteristic (AFROC) curves. The raw curves were transformed from the FROC curves of mass detection on test subset 1 using either the raw images as input and processed with the Laplacian pyramid method (LP) or the GE processed images as input. The FP rate was estimated from the mammograms without masses. The fitted AFROC curves were obtained by applying the ROCKIT program to the transformed AFROC data.

significant, further investigations are needed to study whether these models are valid for analyzing the FROC performance of CAD systems.

The prescreening technique is an important task in a CAD system. A number of researchers have developed methods for detection of suspicious masses on SFMs and CRs. The previous methods produced between 10 to 30 FPs/image for a mass detection sensitivity of approximately 90%. However, it is difficult to compare the effectiveness of the different methods because of the differences in the image recording systems and in the data sets. In this study, we developed a new method that combines gradient field information, which was originally developed for the detection of lung nodules on chest x-ray images,[43] and gray level information[44] for prescreening mass candidates on the FFDMs. The new method produced 18.9 objects/image at 93% sensitivity in the prescreening step, compared with an average of 23.8 objects/image at the same sensitivity by using gradient field information alone.

The texture features in this study were extracted by using the SGLD matrix. A total of 572 features were included in our initial feature pool. These features were also used by our CAD system previously developed for SFMs. An average number of 19.5 features were selected by using a stepwise feature selection method. The A_z values for the LDA classifiers were 0.87 ± 0.02 and 0.88 ± 0.01 on the two training subsets, and 0.89 ± 0.02 and 0.85 ± 0.02 on the test subsets, respectively. The slightly higher test A_z from the first test subset than the A_z from its training subset may indicate that some relatively easy cases were assigned, by chance, to that test set during random partitioning. We also investigated if other features could improve the performance of our CAD system. The different feature spaces that we examined included features extracted from principal component analysis applied to the ROI image, run length statistics texture features extracted from the ROI images, and combination of one or both of these feature spaces with the SGLD feature space. However, the test results showed that a LDA classifier designed in the SGLD feature space alone provided the best performance. Although this was found to be true for both our CAD mass detection system for SFMs developed previously and the current system for FFDMs, it is still difficult to conclude that the SGLD features are the best feature set for classification between breast masses and normal tissues. One major concern of the SGLD feature space is that the dependence of the feature values on the pixel pair distance and angular direction leads to a feature pool with a large number of features. Some features in such a large feature space may provide good performance in classification of masses and normal structures by chance. We attempted to alleviate this problem by using an independent test set to evaluate the classifier performance. However, since we chose the overall system parameters with the knowledge of the performance for the test sets, the evaluation would still amount to validation rather than true testing. We have verified that our CAD system for SFMs can achieve reasonable performance in a true independent data set[36] and a prospective pilot clinical trial.[16] The performance of the current CAD system for FFDMs will have to be evaluated similarly when independent data sets become available.

The detection performance of a CAD system for malignant masses is more important than its performance for all masses. Figures 10(a) and 10(b) indicate that the sensitivity of the system is higher for malignant masses than for benign masses. This is consistent with our observation in previous studies of our CAD system for digitized SFMs.[36] However, since our current data set contained only 23 malignant cases, there will be large statistical uncertainty in the evaluation of sensitivity in this subset. A larger data set is being collected for comparing the detection performances of the CAD system between malignant and benign masses and also for the purpose of classifying malignant and benign masses. Furthermore, CAD algorithms developed for SFMs have been proven to be useful as a second opinion to assist radiologists in mammographic interpretation. Because of the higher SNR and linear response of digital detectors, there is also a potential that FFDMs can improve the sensitivity of breast cancer detection, especially in dense breasts. Several studies have been or are being conducted to compare FFDM with SFM in screening cohorts. It is also important to compare the performance of CAD systems between FFDMs and SFMs. A study is under way to compare the performance of the two systems on pairs of FFDM and SFM obtained from the same patients.[51]

V. CONCLUSION

Several FFDM systems have been approved for clinical applications. It is important to develop CAD systems for breast cancer detection in FFDM. In this work, we developed a CAD system that uses the raw FFDMs as the input. A multiresolution Laplacian pyramid enhancement method was devised to preprocess the raw FFDMs. A new prescreening method that combined gradient field analysis with gray level information was developed to identify mass candidates. Rule-based and LDA classifiers in a feature space which consisted of morphological features and SGLD texture features were designed to differentiate masses from normal tissues. It was found that our CAD system achieved a case-based sensitivity of 70%, 80%, and 90% with an estimate of 0.85, 1.31, and 2.14 FP marks/image, respectively, on normal cases. The results indicate that our mass detection CAD scheme can be useful for detecting masses on FFDMs. Studies are under way to further optimize the processing parameters, the feature extraction, and the classifiers for FP reduction. Comparison of mass detection performance of our CAD system for FFDMs and that for SFMs is also in progress.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA95153, U. S. Army Medical Research and Material Command Grant Nos. DAMD 17-02-1-0214 and W81XWH-04-1-0475. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be

inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC and ROCKIT programs.

[a]Electronic mail: jvwei@umich.edu

[1]American Cancer Society, www.cancer.org 2004, "Statistics for 2004."
[2]C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," Cancer (N.Y.) **75**, 1619–1626 (1995).
[3]S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," Am. J. Roentgenol **171**, 29–33 (1998).
[4]B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," Cancer (N.Y.) **91**, 1699–1703 (2001).
[5]B. J. Hillman, L. L. Fajardo, T. B. Hunter, B. Mockbee, C. E. Cook, R. M. Hagaman, J. C. Bjelland, C. S. Frey, and C. J. Harris, "Mammogram interpretation by physician assistants," Am. J. Roentgenol. **149**, 907–911 (1987).
[6]C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists—Findings from a national sample," Arch. Intern Med. **156**, 209–213 (1996).
[7]R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," Radiology **219**, 192–202 (2001).
[8]J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," J. Natl. Cancer Inst. **95**, 1384–1393 (2003).
[9]F. Shtern, C. Stelling, B. Goldberg, and R. Hawkins, "Novel technologies in breast imaging: National Cancer Institute perspective," Orlando, FL.
[10]C. J. Vyborny, "Can computers help radiologists read mammograms?," Radiology **191**, 315–317 (1994).
[11]H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," Invest. Radiol. **25**, 1102–1110 (1990).
[12]L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology **215**, 554–562 (2000).
[13]T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology **220**, 781–786 (2001).
[14]R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," Am. J. Roentgenology **181**, 687–693 (2003).
[15]S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," Radiology **232**, 578–584 (2004).
[16]M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey *et al.*, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection—A pilot clinical trial," Radiology **231**, 208–214 (2004).
[17]D. Gur, J. H. Sumkin, and H. E. Rockette, "Response to Re: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 1261 (2004).
[18]S. A. Feig, E. A. Sickles, W. P. Evans, and M. N. Linver, "Re. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 1260–1261 (2004).
[19]F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," Radiology **89**, 211–115 (1967).
[20]J. L. Semmlow, A. Shadagopappan, L. V. Ackerman, W. Hand, and F. S. Alcorn, "A fully automated system for screening mammograms," Comput. Biomed. Res. **13**, 350–362 (1980).
[21]T. K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach," Comput. Biomed. Res. **24**, 273–295 (1991).
[22]F. F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," Med. Phys. **18**, 955–963 (1991).
[23]C. Kimme, B. J. O'Laughlin, and J. Sklansky, *Automatic Detection of Suspicious Abnormalities in Breast Radiographs*, Data Structures, Computer Graphics and Pattern Recognition (Academic, New York, 1975).
[24]W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," Radiology **191**, 331–337 (1994).
[25]S. L. Ng and W. F. Bischof, "Automated detection and classification of breast tumors," Comput. Biomed. Res. **25**, 218–237 (1992).
[26]N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms," IEEE Trans. Med. Imaging **15**, 611–619 (1996).
[27]S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," IEEE Trans. Med. Imaging **8**, 377–386 (1989).
[28]D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," IEEE Trans. Med. Imaging **9**, 233–241 (1990).
[29]A. F. Laine, S. Schuler, J. Fan, and W. Huda, "Mammographic feature enhancement by multiscale analysis," IEEE Trans. Med. Imaging **13**, 725–740 (1994).
[30]N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," Med. Phys. **23**, 1685–1696 (1996).
[31]B. Zheng, Y. H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," Acad. Radiol. **2**, 959–966 (1995).
[32]H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computer detection of malignant tumors on digital mammograms," IEEE Trans. Med. Imaging **18**, 369–378 (1999).
[33]L. Li, R. A. Clark, and J. A. Thomas, "Computer-aided diagnosis of masses with full-field digital mammography," Acad. Radiol. **9**, 4–12 (2002).
[34]D. Gur, J. S. Stalder, L. A. Hardesty, B. Zheng, J. H. Sumkin, D. M. Chough, B. E. Shindel, and H. E. Rockette, "Computer-aided detection performance in mammographic examination of masses: Assessment," Radiology **233**, 418–423 (2004).
[35]N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," IEEE Trans. Med. Imaging **15**, 59–67 (1996).
[36]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients," Radiology **224**, 217–224 (2002).
[37]P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **COM-31**, 337–345 (1983).
[38]S. Dippel, M. Stahl, R. Wiemker, and T. Blaffert, "Multiscale constrast enhancement for radiographies: laplacian pyramid versus fast wavelet transform," IEEE Trans. Med. Imaging **21**, 343–353 (2002).
[39]N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1979).
[40]A. Burgess, "On the noise variance of a digital mammography system," Med. Phys. **31**, 1987–1995 (2004).
[41]H. P. Chan, S. C. B. Lo, L. T. Niklason, D. M. Ikeda, and K. L. Lam, "Image compression in digital mammography: Effects on computerized detection of subtle microcalcifications," Med. Phys. **23**, 1325–1336 (1996).
[42]H. Kobatake and S. Hashimoto, "Convergence index filter for vector fields," IEEE Trans. Image Process. **8**, 1029–1038 (1999).
[43]J. Wei, Y. Hagihara, and H. Kobatake, "Detection of rounded opacities on chest radiographs using convergence index filter," ICIAP 99, Venice, 27–29 September, 1999, pp. 757–761.
[44]N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," Med. Phys. **26**, 1642–1654 (1999).

[45]D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," Med. Phys. **24**, 903–914 (1997).
[46]R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Trans. Syst. Man Cybern. **SMC-3**, 610–621 (1973).
[47]C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat. Med. **17**, 1033–1053 (1998).
[48]B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. M. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size.," Med. Phys. **27**, 1509–1522 (2000).
[49]D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," Med. Phys. **16**, 561–568 (1989).
[50]D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," Med. Phys. **31**, 2313–2330 (2004).
[51]J. Wei, B. Sahiner, H. P. Chan, N. Petrick, L. M. Hadjiiski, and M. A. Helvie, "Computer aided diagnosis system for mass detection: Comparison of performance on full-field digital mammograms and digitized film mammograms," RSNA 2003, Chicago, 30 November–5 December 2003, p. 387.

Dual system approach to computer-aided detection of breast masses on mammograms

Jun Wei,[a] Heang-Ping Chan, Berkman Sahiner, Lubomir M. Hadjiiski, Mark A. Helvie, Marilyn A. Roubidoux, Chuan Zhou, and Jun Ge
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109

(Received 13 November 2005; revised 30 August 2006; accepted for publication 31 August 2006; published 18 October 2006)

In this study, our purpose was to improve the performance of our mass detection system by using a new dual system approach which combines a computer-added detection (CAD) system optimized with "average" masses with another CAD system optimized with "subtle" masses. The two single CAD systems have similar image processing steps, which include prescreening, object segmentation, morphological and texture feature extraction, and false positive (FP) reduction by rule-based and linear discriminant analysis (LDA) classifiers. A feed-forward backpropagation artificial neural network was trained to merge the scores from the LDA classifiers in the two single CAD systems and differentiate true masses from normal tissue. For an unknown test mammogram, the two single CAD systems are applied to the image in parallel to detect suspicious objects. A total of three data sets were used for training and testing the systems. The first data set of 230 current mammograms, referred to as the average mass set, was collected from 115 patients. We also collected 264 mammograms, referred to as the subtle mass set, which were one to two years prior to the current exam from these patients. Both the average and the subtle mass sets were partitioned into two independent data sets in a cross validation training and testing scheme. A third data set containing 65 cases with 260 normal mammograms was used to estimate the FP marker rates during testing. When the single CAD system trained on the average mass set was applied to the test set with average masses, the FP marker rates were 2.2, 1.8, and 1.5 per image at the case-based sensitivities of 90%, 85%, and 80%, respectively. With the dual CAD system, the FP marker rates were reduced to 1.2, 0.9, and 0.7 per image, respectively, at the same case-based sensitivities. Statistically significant ($p < 0.05$) improvements on the free response receiver operating characteristic curves were observed when the dual system and the single system were compared using the test sets with either average masses or subtle masses. [DOI: 10.1118/1.2357838]

Key words: computer-aided detection (CAD), mass detection, mammogram, dual system, artificial neural network (ANN)

I. INTRODUCTION

Breast cancer is one of the leading causes of cancer mortality among women.[1] It has been reported that early diagnosis and treatment can significantly improve the chance of survival for patients with breast cancer.[2–4] At present, the most successful method for the early detection of breast cancer is screening mammography.[5] Various methods are being developed to improve the accuracy of breast cancer detection. Double reading by radiologists can reduce the miss rate of radiographic reading. However, double reading will increase the cost of mammographic screening. An alternative method is to use a trained computer-aided detection (CAD) system as a second reader.[6,7] Recent clinical studies have shown that CAD systems are helpful for increasing radiologists' accuracy in detecting breast cancers.[8–13]

A large volume of literature has been published in the CAD area. CAD systems for mammography generally consist of two subsystems: one is a mass detection system and the other is a microcalcification detection system. Detection of masses on mammograms is often more challenging than detection of microcalcifications. The mass detection systems to-date have employed a single-system approach using various techniques for prescreening of mass candidates and classification of true and false positives.[14–24] Our laboratory incorporated two-view mammographic information for improved differentiation of true masses and false positives and obtained promising preliminary results.[22] However, development of new methods to improve the performance of mass detection systems remains an important area of CAD research.

The CAD systems developed so far have mostly used masses seen on current mammograms (i.e., the mammograms on which the masses were detected by radiologists) for training. An important purpose of a CAD system is that it is used as a second reader to alert radiologists to subtle cancers that may be overlooked. To study the ability of a CAD system in detecting subtle cancers that are likely to be missed by radiologists, one way is to evaluate its accuracy in detecting missed cancers on prior mammograms (i.e., the mammograms in previous examinations on which the mass or cancer can be seen retrospectively but was considered

negative or benign at the time of the examination). Some researchers have investigated the performance change of CAD systems when using prior mammograms as input. In our study of mass detection on prior mammograms,[25] we obtained a case-based sensitivity of 74% (20/27) of the malignant masses with 2.2 false positives (FPs) per image. te Brake *et al.*[26] reported that their CAD system has a case-based sensitivity of 34% (22/65) of the cancers which have the appearance of masses or stellate lesions in the prior examinations with 1 FP per image. A commercial system (R2 ImageChecker) also reported detection of 42% (72/172) of the cancers in the prior years which were considered worthy of call-back in retrospect by expert mammographers with about 2 FP marks/case.[27] Zheng *et al.*[23] reported that their CAD system trained with current mammograms could not perform optimally in prior mammograms and vice versa; whereas the same system trained with prior mammograms can perform better on detecting the masses on prior mammograms. Recently, an assessment study[28] was conducted to compare the performance of two commercial systems and one research CAD system on current mammograms and prior mammograms. The results showed that the true positive (TP) fraction for CAD systems on prior mammograms of 39 breasts with malignant masses ranged from 15% to 26% with 0.28 to 0.41 FP marks/image. Although the detection performance reported in the different studies vary, probably due to the differences in the data set used, these studies indicate that the sensitivities of current CAD systems in detecting subtle masses on prior mammograms are substantially lower than that obtained from detection on current mammograms. The difficulty in recognizing the subtle and possibly different features of the masses on priors compared to those of the masses on current mammograms may be one of the factors that causes oversight for both radiologists and the CAD systems.

The goal of pattern recognition is to achieve the best possible classification performance in the task at hand. Researchers had shown that, for a class of objects with a wide range of characteristics, the classification performance can be improved by using combination of classifiers whereby objects of certain characteristics are classified by one classifier using a set of features and objects of different characteristics by another classification scheme based on different features.[29–35] The advantage of using combination of classifiers is that it may stabilize the training of classifiers even with a relatively small sample size because each classifier does not have to accommodate a wide range of characteristics and features.[36,37] These observations motivated our interest in the design of a dual CAD system for mass detection.

Since the missed cancers on prior mammograms represent the difficult cases that are more likely to be missed by radiologists if similar cancers occur on screening mammograms, it is important to improve the sensitivity of the CAD system in detecting these cancers. On the other hand, when a CAD system is applied to a new mammogram in clinical practice, it has to detect breast lesions of all degrees of subtlety effectively. However, it is difficult to train a single CAD system to provide optimal detection for all lesions over the entire spectrum of subtlety because the classifiers have to make compromises to accommodate cancers of a wide range of characteristics. Therefore, we have been exploring a new dual CAD system approach that combines a CAD system trained with retrospectively seen masses on prior mammograms with a CAD system trained with masses detected on current mammograms.[38,39] In this paper, we will describe the design of the dual CAD system and report our current results.

II. MATERIALS AND METHOD

A. Data sets

All mammograms in this study were collected from patient files in the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. The mammograms were digitized with a LUMISYS 85 laser film scanner with a pixel size of 50 μm $\times$ 50 μm and 4096 gray levels. The scanner was calibrated to have a linear relationship between gray levels and optical densities (O.D.) from 0.1 to greater than 3 O.D. units. The nominal O.D. range of the scanner is 0–4. The full resolution mammograms were first smoothed with a 2×2 box filter and subsampled by a factor of 2, resulting in 100 μm $\times$ 100 μm images. The images at a pixel size of 100 μm $\times$ 100 μm were used for the input of our CAD system.

We collected three data sets. The first data set contained 115 cases with confirmed masses. Each case included the current mammograms that prompted the radiologist to work up the mass. This is referred to as the "average" mass set. All of the cases in the average mass set had two mammographic views: the craniocaudal view and the mediolateral oblique view or the lateral view, thus yielding a total of 230 mammograms. There were 115 masses (67 malignant masses and 48 benign masses) in this data set, of which 105 were biopsy-proven and 10 were determined to be benign by long-term follow-up.

The second data set was composed of the prior mammograms dated one to two years earlier than the mammograms of the same patients in the average mass set. Since the masses on prior mammograms are on average subtler than those on current mammograms, this data set is referred to as the "subtle" mass set. On 5 of the 115 patients, no mass or focal density could be identified on either view of the prior mammograms. Therefore, the subtle mass set was composed of 110 cases (62 malignant and 48 benign). For the purpose of training the subtle mass detection system, the subtle masses do not have to be obtained from the same cases as the average mass set but we used the available prior mammograms for these mass cases in our database. Nineteen of the 110 cases had two prior mammogram examinations. Of the 129 examinations in the subtle mass set, 123 had two mammographic views and 6 had three views, with a total of 264 mammograms. Many of the subtle masses on the prior mammograms could be identified only as a focal density corresponding to the location of the subsequently detected mass on the current mammograms. On 44 of the two-view prior

TABLE I. Description of cases in the average and subtle mass data sets and the subsets for training and testing in the cross-validation scheme.

	Mass subset 1		Mass subset 2	
	Average mass subset	Subtle mass subset	Average mass subset	Subtle mass subset
Total No. of cases	57	54	58	56
Cases with two prior examinations	NA	10	NA	9
Exams with two views	57	58	58	65
Exams with three views	0	6	0	0
Total No. of images	114	134	116	130
No. of negative images	0	25	0	19
No. of mass images for training	114	109	116	111
No. of two-view pairs for testing	57	64	58	65
No. of images for testing	114	128	116	130
No. of malignant masses	36	33	31	29
No. of benign masses	21	21	27	27

mammograms, the mass location was evident only on one view. Table I summarizes the information for the average and subtle mass subsets.

The third data set was composed of 260 normal bilateral two-view mammograms obtained from 65 patients. No masses were evident on these mammograms upon review by the experienced radiologist.

The two mass data sets were used to estimate the detection sensitivity and the normal data set was used for estimating the FP marker rate. For the mass data sets, the true locations of the masses were identified by an experienced MQSA radiologist using all available imaging and clinical information. The radiologist also provided an estimate of the longest diameter of the mass, descriptors of its margin and shape, a visibility rating, and an estimate of the breast density in terms of BI-RADS category. Figure 1 shows the distributions of mass sizes, mass shapes, mass margins, and their visibility on a 10-point rating scale with 1 representing the most visible masses and 10 the most difficult case relative to the cases seen in their clinical practice. The masses had a mean of 13.7 mm and a median of 12 mm in the average data set and a mean of 9.7 mm and a median of 10 mm in the subtle data set. Figure 2 shows the breast density for both the normal data set and the mass data sets. As can be seen from the distributions of the mass characteristics, the average masses on the current mammograms and the subtle masses on the priors had large overlap. Nevertheless, on average, the subtle masses were smaller in size and less conspicuous on the mammograms.

B. Methods

In order to improve the sensitivity of detecting breast lesions of all degrees of subtlety, we developed a new dual system approach which combines a system trained with average masses with another system trained with subtle masses. When the trained dual system is applied to an unknown mammogram, the two CAD systems are used in parallel to detect suspicious objects on a single mammogram. No prior mammogram is needed. The additional FPs from the use of the two systems are reduced by an information fusion stage. We will refer to the two systems separately trained with the average masses and the subtle masses as "single" CAD systems in the following discussions.

We randomly separated the mass data sets by case into two independent subsets. Both the average and subtle mass subsets followed the same case grouping so that mammograms from the same case would not be separated into the training subset for one single CAD system and the test subset for the other single CAD system in a cross-validation cycle. Table I shows the subsets of cases in the average and subtle mass data sets. Two-fold cross validation was used for training and testing the algorithms. The training included selecting proper parameters for each single CAD system and for information fusion. Once the training with one mass subset was completed, the parameters were fixed for testing with the other mass subset. The training and test mass subsets were switched and the training and test processes were repeated. The CAD systems were trained with single mammograms. To maximize the number of training images with masses, all images with a visible mass were included regardless of whether they were a part of a two-view or three-view case when the subtle mass subset was used as a training set. However, when the subtle mass subset was used as a test set, only two views were included for each case because we used two-view mammograms to derive the case-based test performance. For cases containing three views, we therefore included only two of the views in testing. We also included cases with the mass visible on only one of the two views. After the two-fold cross validation testing, the overall detection performance was evaluated by combining the performances of the two test subsets. The trained algorithms with the fixed parameters were also applied to the normal set of mammograms, which was not used during training, to estimate the FP rate in screening mammograms.

1. Single CAD system overview

The major steps in the two single mass detection systems are similar but the feature spaces and classifiers for FP reduction in each system were designed separately to suit the characteristics of average and subtle masses, respectively. The two systems are therefore described together in the following but the differences will be pointed out whenever applicable. Each single CAD system consists of four processing steps: (1) prescreening of mass candidates, (2) segmentation of suspicious objects, (3) feature extraction and

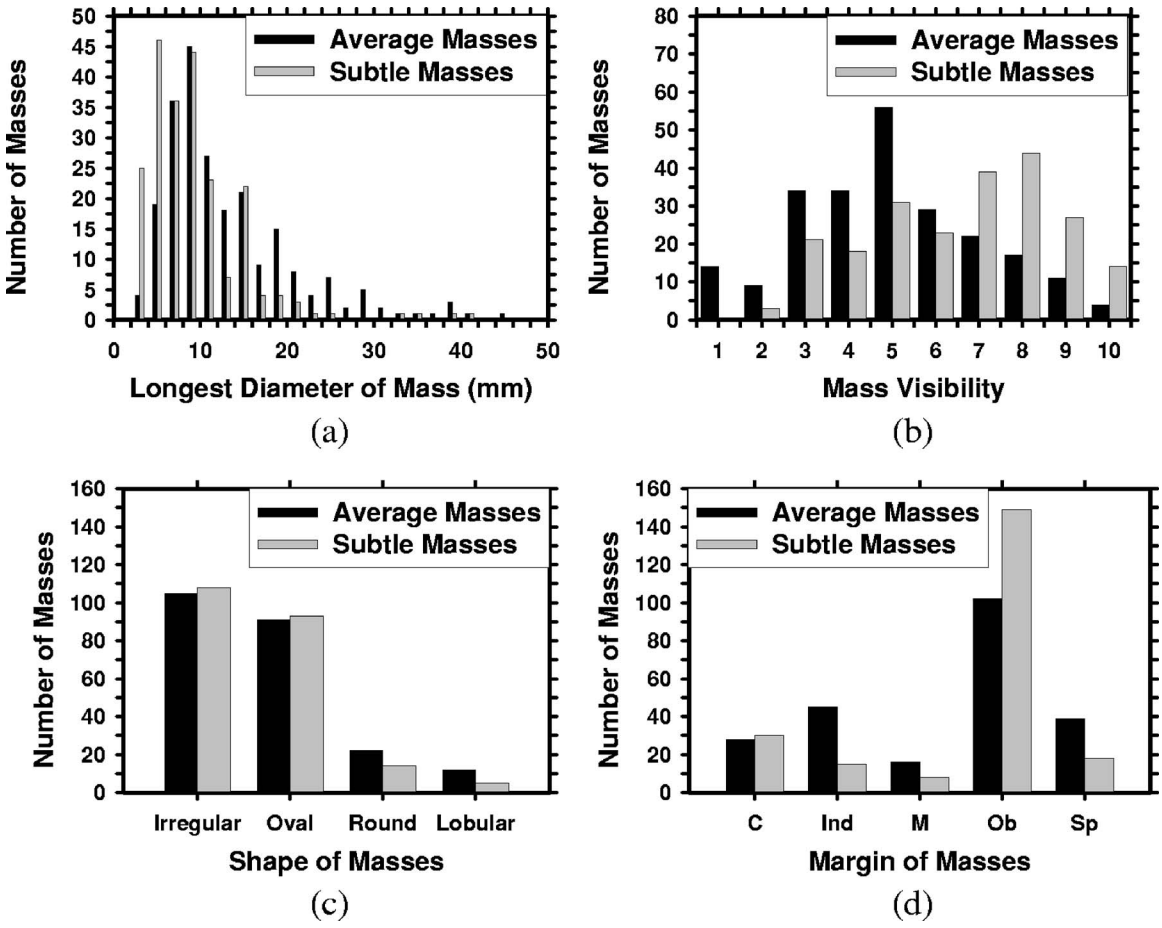


FIG. 1. The characteristics of the masses in our mass data set: (a) distribution of mass sizes, (b) distribution of mass visibility on a 10-point rating scale with 1 representing the most visible masses and 10 the most subtle masses relative to the cases seen in clinical practice, (c) distribution of mass shapes, (d) distribution of mass margins, C: circumscribed, Ind: indistinct, M: microlobulated, Ob: obscured, Sp: spiculated.

analysis, and (4) FP reduction by classification of normal tissue structures and masses. The block diagram for the detection scheme is shown in Fig. 3.

For the prescreening stage, we have developed a two-stage gradient field analysis method which not only uses the shape information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis.[24,40] Locations of high radial gradient convergence are labeled as mass candidates. After prescreening, the suspicious objects are identified by using a two-stage segmentation method.[41] First, the background-

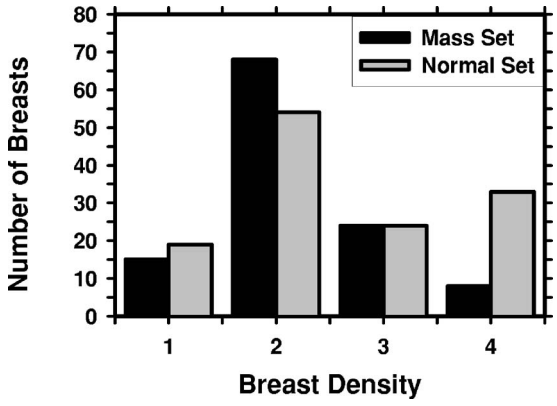


FIG. 2. The distribution of breast density in terms of BI-RADS categories estimated by an MQSA radiologist.

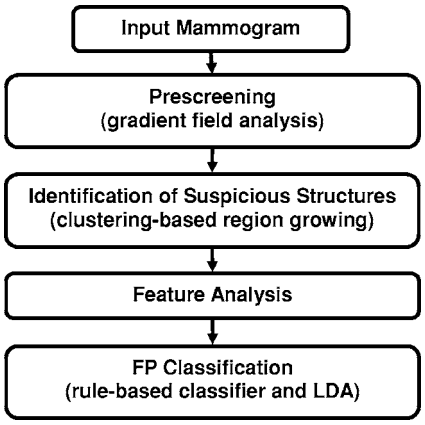


FIG. 3. Schematic diagram of our single CAD system for mass detection. The FP classification stage includes rule-based classification, a morphological LDA classifier, and a texture feature LDA classifier for differentiating masses from normal breast tissues.

corrected ROI is weighted by a two-dimensional Gaussian function with $\sigma=256$ pixels to enhance the central region. Sobel filtering is then applied to the Gaussian-weighted ROI to generate another enhanced image. Second, a *k*-means clustering using the pixel values from these two images as features is used to segment the object. For each suspicious object, eleven morphological features[21] were extracted. Rule-based and linear discriminant classifiers were trained by using the training data set only to remove the detected structures that were substantially different from breast masses. For the system trained with average masses, global and local multiresolution texture analysis[42] were performed in each ROI by using the spatial gray level dependence (SGLD) matrices. A total of 364 features were extracted from global texture analysis. Local texture features were extracted from the local region containing the detected object and the peripheral regions within each ROI. A total of 208 features were extracted for local texture analysis. For the system trained with subtle masses, instead of the SGLD texture features, gray level features and run length statistics analysis (RLS) texture features[43] were extracted inside and outside of each mass region on the original image and gradient field image. The gray level features included the contrast of the object relative to the surrounding background, the minimum and the maximum gray levels, and the characteristics derived from the gray level histogram in the regions inside and outside of each object including skewness, kurtosis, energy, and entropy. Five RLS texture features were extracted in both the horizontal and vertical directions: short runs emphasis, long runs emphasis, gray level nonuniformity, run length nonuniformity, and run percentage. A total of 66 features were extracted for the system trained with subtle masses.

In order to obtain the best texture feature subset and also reduce the dimensionality of the feature space to design an effective classifier, stepwise feature selection with linear discriminant analysis (LDA) was applied to the training subset. The detailed procedure has been described elsewhere.[24,44,45] Briefly, at each step one feature was entered or removed from the feature pool by analyzing its effect on the selection criterion, which was chosen to be the Wilks' lambda in this study. Since the appropriate values of thresholds for feature entry, feature elimination, and tolerance of correlation for feature selection were unknown, we used an automated simplex optimization method to search for the best combination of thresholds in the parameter space. The simplex algorithm used a leave-one-case-out resampling method within the training subset to select features and estimate the weights for the LDA classifier. To have a figure-of-merit to guide feature selection, the test discriminant scores from the left-out cases were analyzed using receiver operating characteristic (ROC) methodology.[46] The accuracy for classification of masses and FPs was evaluated as the area under the ROC curve, A_z. In this approach, feature selection was performed without the left-out case so that the test performance would be less optimistically biased.[47] However, the selected feature set in each leave-one-case-out cycle could be slightly different because every cycle had one training case different from the other cycles. In order to obtain a single trained classifier to apply to the independent test subset, a final stepwise feature selection was performed with the best combination of thresholds, found in the simplex optimization procedure, on the entire training subset to obtain the final set of features and estimate the weights of the LDA. Note that the entire process of feature selection and classifier weight estimation was performed within the training subset. The LDA classifier with the selected feature set was then fixed and applied to the independent test subset. The training and testing processes were performed independently for the two-fold cross-validation sets.

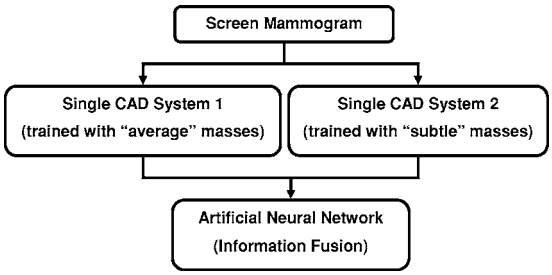


FIG. 4. Schematic diagram of proposed dual CAD system for mass detection. BP-ANN is used for information fusion.

2. Training and test for dual system

The block diagram for the dual system is shown in Fig. 4. During the training of the dual system, we used the current and prior mammograms from the same patients. The current mammograms that contained the average masses were only used to train the first single CAD system. The prior mammograms that contained the subtle masses were only used to train the second single CAD system. The prescreening and the segmentation steps in the two systems are identical. Since the morphological appearances of average and subtle masses are different, the rules in the morphological rule-based FP classification are trained differently for the two single CAD systems. During testing with an independent mammogram, the dual system keeps all the suspicious objects that satisfy the FP classification rules of either single CAD system and applies the LDA classifiers from both single systems to each object. Each object thus has two LDA scores.

To merge the information from the two CAD systems, a fusion scheme was developed for our dual system. In this study, a feed-forward backpropagation artificial neural network (BP-ANN) was trained to classify the masses from normal tissues by combining the output information from the two single CAD systems. The LDA classifiers from the two single CAD systems were applied to each detected object. The two LDA discriminant scores for each object were used as input to the BP-ANN. The BP-ANN had an input layer with two nodes, a hidden layer with N nodes, and an output layer with one node. The nodes were interconnected by weights and information propagated from one layer to the next through a log-sigmoidal activation function. The learning of the ANN was a supervised process in which known training cases were input to the ANN. The performance func-

tion for the network was the mean-squared error between the network outputs and the target outputs. The weights of the network were adjusted iteratively by a feedforward backpropagation procedure to minimize the error. Detailed description of the backpropagation neural network can be found in the literature.[48,49]

To choose the number of hidden nodes (N) in the BP-ANN, we used a three-fold cross-validation method within the training subset. We randomly separated the entire training subset including all detected objects into three independent groups. The objects belonging to the same case were separated into the same group. For a given N, three training cycles were performed, in each of which two of the three groups were used to train the BP-ANN and the left-out group was used to test its performance. The A_z value obtained from the ANN output scores for the test group was used as the performance index for that training cycle. The average of the A_z values from the three test groups represented the performance of the BP-ANN with N hidden nodes. In our experiment, a BP-ANN with 3 hidden nodes provided the largest average A_z value and was therefore chosen. The weights of the chosen BP-ANN were retrained with the entire training subset. The BP-ANN with the trained weights was used to merge the information from the two single CAD systems.

To test the dual system, the two trained single CAD systems, one trained with the average mass set and the other with the subtle mass set, were applied in parallel to each single "unknown" mammogram in the independent test subset. No prior mammogram was needed during testing.

3. Evaluation methods

The detected individual objects were compared with the "truth" ROI marked by the experienced radiologist, as described earlier. A detected object was scored as TP if the overlap between the bounding box of the detected object and the bounding box of the true mass relative to the larger of the two bounding boxes was over 25%. Otherwise, it would be scored as FP. The 25% threshold was selected as described in our previous study.[21]

The FP marker rate was estimated in two ways: one from detection on the same test subsets with masses, the other from detection on the normal data set of negative mammograms. For the latter, we applied the trained dual CAD system to the normal data set. The number of FP marks produced by the CAD system was determined by counting the detected objects on the normal cases. The mass detection sensitivity was determined by counting the detected masses on the test mass subset. The detection performance of the CAD system was assessed by free response ROC (FROC) analysis. A FROC curve was obtained by plotting the mass detection sensitivity as a function of FP marks per image either obtained from the mass data subset or the normal set at the corresponding decision threshold.

FROC curves were presented on a per-mammogram and a per-case basis. For image-based FROC analysis, the mass on each mammogram was considered an independent true object. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either or both masses on the two views was considered to be a TP detection.

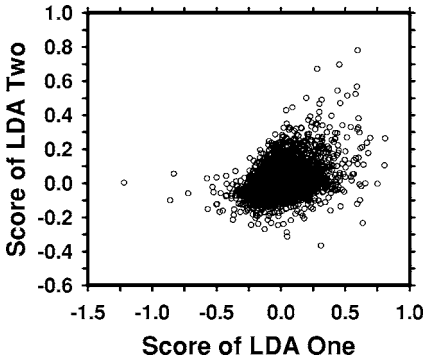


FIG. 5. An example of a scatter plot of the LDA scores from the two single CAD systems which are used as input to the BP-ANN. The correlation coefficient between the scores of two LDA classifiers is 0.46, indicating that the two LDA scores are essentially independent features.

Since we used two-fold cross validation method for training and testing, we obtained two test FROC curves, one for each test subset, for each of the conditions (e.g., single CAD system approach or dual system approach). To summarize the results for comparison, an average test FROC curve was derived by averaging the FP rates at the same sensitivity along the FROC curves of the two corresponding test subsets.

In order to compare the performance of the single CAD system and the dual CAD system, we applied the alternative free-response ROC (AFROC) method and the jackknife free-response ROC (JAFROC) method developed by Chakraborty *et al.*[50,51] to the pairs of FROC curves. In the AFROC method, the FROC data are first transformed by counting the number of false-positive images (FPIs) instead of the FPs per image. The confidence rating of a FPI is determined by the highest confidence FP decision on the image regardless of how many lower confidence FP decisions are made on the same image. The ROCKIT curve fitting software and statistical significance tests for ROC analysis developed by Metz *et al.*[46] can then be used to analyze the AFROC data.

III. RESULTS

Figure 5 shows an example of the two-dimensional feature space that was used as the input to the BP-ANN being trained to merge the information from the two single CAD subsystems. The two features are the output scores of the LDA classifiers trained with the average masses and with the subtle masses. The correlation coefficients of the two features are 0.46 and 0.44 for each of the training subsets, respectively. The low correlation indicated that the two single CAD systems extracted relatively independent features from the object. The A_z values of the chosen ANN were 0.92 ± 0.01 and 0.87 ± 0.01, respectively, as estimated by validation in the training process. The ANN classifiers achieved A_z values

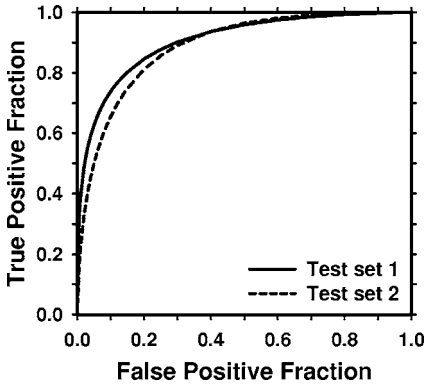


FIG. 6. The test ROC curves for the BP-ANN classifiers from the two independent mass subsets. The ANN classifiers achieved an A_z value of 0.90 ± 0.02 for test subset 1 and 0.89 ± 0.01 for test subset 2 in the classification of mass and normal breast tissues.

of 0.90 ± 0.02 and 0.89 ± 0.01 on the two independent test subsets, respectively. Figure 6 shows the ROC curves for the two test subsets.

In order to evaluate the effectiveness of our dual system approach, we compared its performance on the test subsets containing average masses with two other single CAD systems: the CAD system trained only on the average mass set and the CAD system trained on both the average and the subtle mass sets. When a single CAD system was trained only with the average masses, the number of selected features was 21 (14 global and 7 local) and 16 (10 global and 6 local) texture features for the two independent training subsets, respectively. When the CAD system was trained with both the average and the subtle masses, the number of selected features was 17 (11 global and 6 local) and 18 (7 global and 11 local) texture features for the two independent training subsets, respectively.

For the dual system, the single system trained with the average masses was the same as that described earlier. For the single system trained with subtle masses, four (2 gray level and 2 RLS texture) and five (3 gray level and 2 RLS texture) features were selected for the two independent training subsets, respectively.

The average test FROC curves of the dual CAD system on the test subsets with average masses were compared to those of the single CAD systems in Fig. 7. The FP rates were estimated from the mass data set. The dual CAD system achieved a case-based sensitivity of 80%, 85%, and 90% at 0.6, 0.8, and 1.0 FPs/image, respectively, compared with 1.3, 1.5, and 1.8 FPs/image on the single CAD system trained with average masses alone. The performance of the single CAD system trained with both the average masses and the subtle masses was comparable to that trained with average masses alone, with FP rates of 1.4, 1.6, and 1.8 FPs/image at the same sensitivities, respectively. Figure 8 shows the comparison of the three average test FROC curves, similar to those shown in Fig. 7, except that the FP rates were estimated from the normal data set. The FP rates at a few selected sensitivities for the dual and single CAD systems were summarized in Table II.

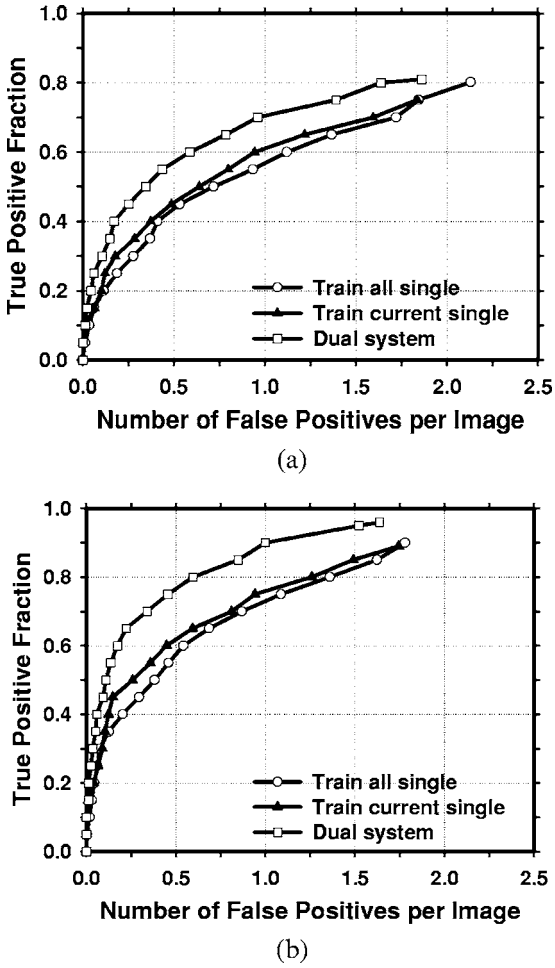


FIG. 7. Comparison of the average test FROC curves obtained from averaging the FROC curves of the two independent average-mass subsets. Three CAD systems were compared: a single CAD system trained with average masses alone, a single CAD system trained with both the average and the subtle masses, and the dual CAD system. The FP rate was estimated from the mammograms with masses. (a) Image-based FROC curves, (b) case-based FROC curves.

In this study, we have 67 malignant cases in the average mass set. Figure 9 compares the average test FROC curves of the single CAD system and the dual system for detection of malignant masses. The result for the single CAD system trained with average masses was shown and the FP rate was estimated from the mammograms without masses. In this case, the dual CAD system achieved a case-based sensitivity of 80%, 85%, and 90% at 0.6, 0.9, and 1.2 FP marks/image, respectively, compared with 1.1, 1.6, and 2.0 FP marks/image on the single CAD system.

An important purpose of a CAD system is to serve as a second reader to alert radiologists to subtle cancers that may be overlooked. Figures 10 and 11 compare the average FROC curves of the single CAD system and the dual system for detection in the test subsets with subtle masses. The TP rate in Fig. 10 was estimated by including both malignant and benign masses and that in Fig. 11 was estimated from malignant masses only. The single CAD system trained with average masses alone was used. The FP rates for both sys-

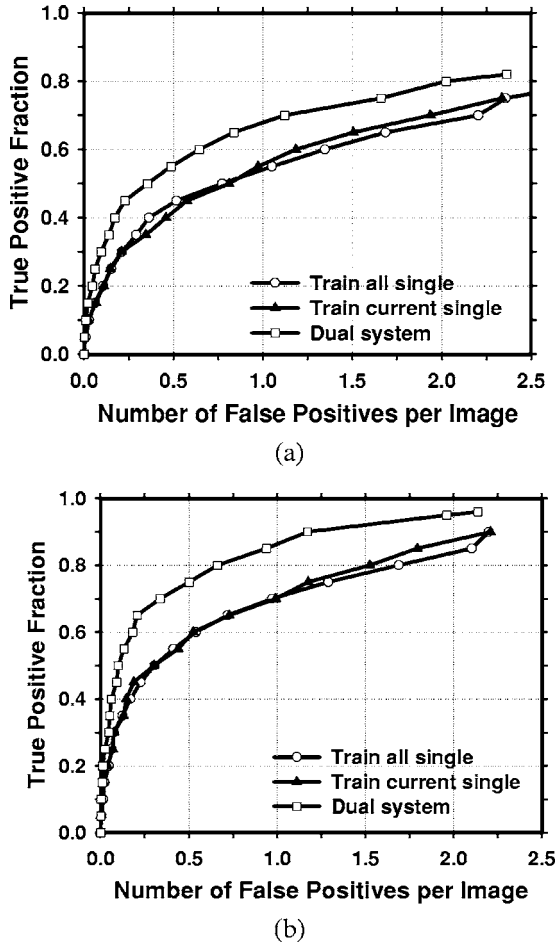


FIG. 8. Comparison of the average test FROC curves obtained from averaging the FROC curves of the two independent average-mass subsets. Three CAD systems were compared: a single CAD system trained with average masses only, a single CAD system trained with the average and the subtle masses, and the dual CAD system. The FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) case-based FROC curves.

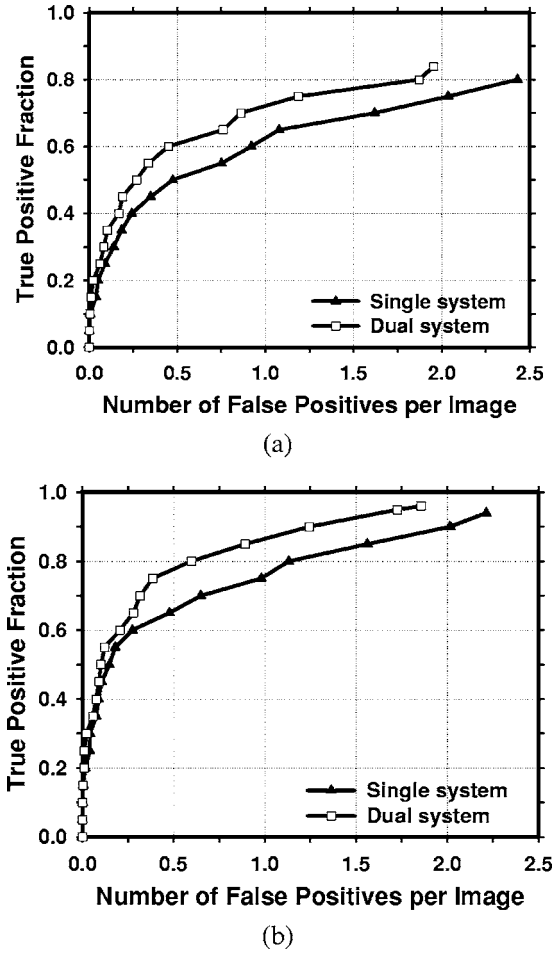


FIG. 9. Comparison of the average test FROC curves of the single CAD system and the dual CAD system for detection of malignant masses in the average data set. The single system trained with average masses alone was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) case-based FROC curves.

tems were estimated from the mammograms without masses. The dual CAD system achieved a case-based sensitivity of 50% at 0.7 FP marks/image for all masses and at 0.5 FP marks/image for malignant masses only, compared with 1.4 FP marks/image for all masses and 1.1 FP marks/image for malignant masses only using the single CAD system.

TABLE II. Comparison of case-based detection performance between the dual system and the single CAD system trained with average masses alone. The FP marker rates were estimated from detection on the normal data set. The FROC curves were obtained by averaging the FROC curves of the two test subsets.

	Average mass test set (FP marks/image)		Subtle mass test set (FP marks/image)	
TP	Single system	Dual system	Single system	Dual system
90%	2.2	1.2		
80%	1.5	0.7		2.8
70%	1.0	0.3	2.4	2.3
60%	0.5	0.2	1.8	1.5
50%	0.3	0.1	1.4	0.7

Table II summarizes the test results on the average and subtle mass sets for the dual system and the single CAD system trained with average masses at different sensitivity levels. The FP marker rates were estimated from the detection on the normal data set.

The comparison of the FROC curves for the dual CAD system and the single CAD system in terms of the area under the fitted AFROC curve (A_1) and the p values for both test subsets with average masses was summarized in Table III. The differences between the A_1 values for the two systems were statistically significant ($p<0.05$). The fitted AFROC curves, however, did not fit very well to the transformed AFROC data, as we discussed previously.[24] For the JAFROC method, Chakraborty *et al.* provided software to estimate the statistical significance of the difference between two FROC curves. The comparison of the figure-of-merit (FOM) and the p values was also summarized in Table III. The differences

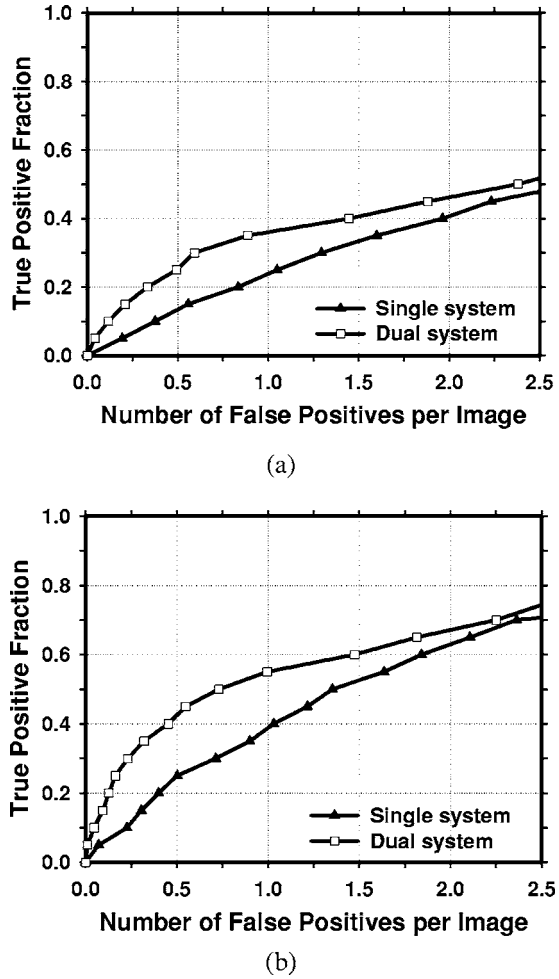


FIG. 10. Comparison of the average test FROC curves for the single CAD system and the dual CAD system for detection of the subtle masses on the prior mammograms. The single CAD system trained with average masses alone was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) case-based FROC curves.

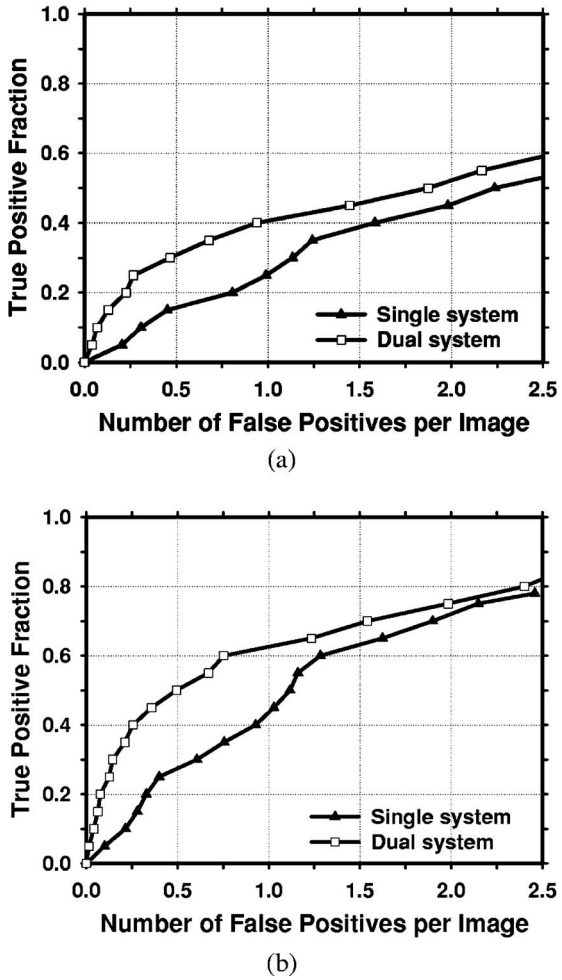


FIG. 11. Comparison of the average test FROC curves for the single CAD system and the dual CAD system for detection of subtle malignant masses on the prior mammograms. The single CAD system trained with average masses alone was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) case-based FROC curves.

between the FOM of the dual CAD system and that of the single CAD system for both test subsets were again statistically significant ($p < 0.05$).

The comparison of A_1, the FOM, and the p values for the dual system and the single system trained with average masses in detecting subtle masses was summarized in Table IV. It was found that the differences between the results of the dual CAD system and those of the single CAD system on the two test subsets containing subtle masses were statistically significant by both the JAFROC and the AFROC methods.

IV. DISCUSSION

The masses on prior mammograms are more subtle and more difficult to detect than the masses on current mammograms. In this study, we developed a dual CAD system, which combines a system trained with masses on prior mammograms and a system trained with masses detected on current mammograms. We have demonstrated that this dual system can increase the accuracy of detecting both average masses and subtle masses. The comparisons of the dual system with that of the single CAD system trained with average masses alone and that of the single CAD system trained with both average and subtle masses (Fig. 7) indicate that the gain in the detection accuracy of the dual system could not be achieved by simply using a larger training set with both average and subtle masses. In fact, it is interesting to note that the performance of the single CAD system trained with both the average and the subtle masses appeared to be degraded slightly, in comparison with the single system trained with average masses alone, when it was applied to the test set of average masses. The decreased performance may reflect the compromise made when the single CAD system was trained to accommodate a wide range of lesion characteristics. Thus, the dual system approach may have improved its performance through other factors, including the flexibility in using different feature spaces and training the parameters for each type of masses and the information fusion combining the two single CAD systems effectively.

TABLE III. Estimation of the statistical significance in the difference between the FROC performance of the dual system and the single CAD system trained with average masses alone when the systems were evaluated on the average mass test subsets. The FROC curves with the FP marker rates obtained from the normal data set were compared.

	A_1 (AFROC)				FOM (JAFROC)			
	All cases		Malignant cases		All cases		Malignant cases	
	Test subset 1	Test subset 2	Test subset 1	Test subset 2	Test subset 1	Test subset 2	Test subset 1	Test subset 2
Single system	0.45	0.44	0.47	0.52	0.48	0.48	0.53	0.55
Dual system	0.55	0.53	0.58	0.62	0.60	0.56	0.63	0.64
p values	0.0004	0.0156	0.0003	0.0318	<0.0001	0.007	0.0004	0.0252

For the comparison of the different systems, we analyzed the false negatives (FNs) of the single CAD systems and the dual CAD system when the test subsets with average masses were used. It was found that the FN rates of the single CAD system trained with average masses, the single CAD system trained with subtle masses, and the dual system were 23.9% (55/230), 28.3% (65/230), and 16.5% (38/230), respectively, after FP reduction by the morphological LDA classifier in each system. Twenty-nine masses were missed by both of the single systems. By using the dual system, 53 masses that were FNs for either single system could be detected. However, the masses that were missed by both of the single CAD systems could not be recovered by the dual CAD system.

Our motivation of this study is to improve the performance of a CAD system for mass detection. A CAD detection system is generally intended for use in screening mammography. At the screening stage, all lesions of concern should be pointed out to radiologists so that the radiologists can judge if a recall is warranted. If a detection system is trained to mark only the malignant lesions, it may be attempting to play the role of a triage system (alerting radiologists to work up only "malignant" cases) rather than that of a second reader. Furthermore, since computerized lesion detection or characterization on mammograms is not 100% sensitive, it will be confusing to the radiologists whether an unmarked suspicious lesion is missed or it is considered benign by the computer. We believe that computer-aided diagnosis (CADx) may be used in different ways in conjunction with a CAD detection system, for example, the likelihood of malignancy may be estimated by the CADx system and displayed for every detected lesion, and/or a CADx system may be used during diagnostic workup. Either way the CAD system will first alert radiologists to all masses, leaving the assessment of malignancy or benignity to a second stage and with the radiologist being the primary decision maker. The training set thus included both malignant and benign masses.

For a CAD system, its performance for detecting malignant masses is more important than its performance for detecting all masses. The FROC curves for detection of malignant masses on the average data set and the subtle data set, shown in Figs. 9 and 11, respectively, indicated that the dual system could also achieve an improvement in the detection performance over that of the single system. The differences in the A_1 and the FOM for the detection of malignant cases in the average and subtle mass test subsets were statistically significant, as shown in Tables III and IV, respectively.

In screening mammography, the cancer rate is 3–5 per 1000. Most of the mammograms are normal. Therefore, some CAD researchers and users estimate the FP rate using

TABLE IV. Estimation of the statistical significance in the difference between the FROC performance of the dual system and the single CAD system trained with average masses alone when the systems were evaluated on the subtle mass test subsets. The FROC curves with the FP marker rates obtained from the normal data set were compared.

	A_1 (AFROC)				FOM (JAFROC)			
	All cases		Malignant cases		All cases		Malignant cases	
	Test subset 1	Test subset 2	Test subset 1	Test subset 2	Test subset 1	Test subset 2	Test subset 1	Test subset 2
Single system	0.17	0.20	0.24	0.25	0.21	0.23	0.24	0.26
Dual system	0.28	0.25	0.35	0.34	0.30	0.28	0.36	0.34
p values	<0.0001	0.046	<0.0001	0.0067	0.0007	0.048	<0.0001	0.0035

normal mammograms[52–54] because it reflects how the CAD system performs in terms of specificity and whether the CAD system may cause extra efforts for radiologists to double check the marked locations or unnecessary recalls in a screening setting. Furthermore, for CAD systems that set a maximum number of detected objects at the output, estimating the number of FPs using images with lesions can potentially lead to an optimistic bias for the FROC curve because one of the detected objects will likely be the true lesion. The FP rate can thus be underestimated by as much as 1 per image. In addition, the JAFROC analysis requires that the FP rates be estimated on normal images. We therefore reported the FP rates of our CAD systems on both mammograms with masses and without masses to facilitate comparison with other CAD systems in case investigators may evaluate their FP rates in either way.

In this study, we evaluated the performance of the trained CAD systems with an independent test set using the two-fold cross validation method. Although the selection of parameters and features was performed using the training set, we had full knowledge of the performance for the test set so that the selections could be optimistically biased. True independent testing will have to be performed with unknown cases that have never been used for testing the CAD system before, such as those in a prospective clinical trial. However, this test step is beyond the scope of our current developmental process. Since we used the same cross-validation method for evaluation of the dual system and the single CAD systems, the comparison of their relative performances is expected to be less biased than their individual performances.

V. CONCLUSION

We have proposed a new dual system approach which combines a system trained with subtle masses on prior mammograms and a system trained with average masses on current mammograms. The dual system achieved higher sensitivities at the corresponding FP rates than a single CAD system trained with average masses alone or trained with both average masses and subtle masses. Alternatively, the dual system had lower FP rates than the single CAD system at corresponding sensitivities. The improvement in the FROC curves by the dual system approach was found to be statistically significant ($p < 0.05$) for both average masses and subtle masses using either the AFROC or the JAFROC method. Our results indicate that the dual system approach is promising for improving the performance of CAD systems for mass detection on mammograms.

ACKNOWLEDGMENTS

This work is supported by U. S. Army Medical Research and Materiel Command Grant Nos. W81XWH-1-04-1-0475, DAMD 17-02-1-0214, and USPHS Grant No. CA95153. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program and to Dev Chakraborty, Ph.D., for the JAFROC program.

[a]Electronic mail: jvwei@umich.edu

[1]American Cancer Society, www.cancer.org, "Statistics for 2005."
[2]C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," Cancer **75**, 1619–1626 (1995).
[3]S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," AJR, Am. J. Roentgenol. **171**, 29–33 (1998).
[4]B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," Cancer **91**, 1699–1703 (2001).
[5]H. C. Zuckerman, "The role of mammography in the diagnosis of breast cancer," in *Breast Cancer, Diagnosis and Treatment*, edited by I. M. Ariel and J. B. Cleary (McGraw-Hill, New York, 1987).
[6]F. Shtern, C. Stelling, B. Goldberg, and R. Hawkins, "Novel technologies in breast imaging: National Cancer Institute perspective," Orlando, FL.
[7]C. J. Vyborny, "Can computers help radiologists read mammograms?," Radiology **191**, 315–317 (1994).
[8]T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology **220**, 781–786 (2001).
[9]M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey *et al.*, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection—A pilot clinical trial," Radiology **231**, 208–214 (2004).
[10]R. L. Birdwell, P. Bandodkar, and D. M. Ikeda, "Computer-aided detection with screening mammography in a university hospital setting," Radiology **236**, 451–457 (2005).
[11]D. Gur, J. H. Sumkin, H. E. Rockette, M. A. Ganott, C. Hakim, L. A. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 185–190 (2004).
[12]S. A. Feig, E. A. Sickles, W. P. Evans, and M. N. Linver, "Re. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 1260–1261 (2004).
[13]T. E. Cupples, "Impact of computer-aided detection (CAD) in a regional screening mammography program," Radiology **221**(P), 520 (2001).
[14]S. L. Ng and W. F. Bischof, "Automated detection and classification of breast tumors," Comput. Biomed. Res. **25**, 218–237 (1992).
[15]N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," Med. Phys. **23**, 1685–1696 (1996).
[16]B. Zheng, Y. H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," Acad. Radiol. **2**, 959–966 (1995).
[17]N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms," IEEE Trans. Med. Imaging **15**, 611–619 (1996).
[18]H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," IEEE Trans. Med. Imaging **15**, 235–245 (1996).
[19]Z. M. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," Acad. Radiol. **5**, 155–168 (1998).
[20]W. Qian, L. H. Li, and L. P. Clarke, "Image feature extraction for mass detection in digital mammography: Influence of wavelet analysis," Med. Phys. **26**, 402–408 (1999).
[21]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients," Radiology **224**, 217–224 (2002).
[22]S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Improvement of computerized mass detection on mammograms: Fusion

of two-view information," Med. Phys. **29**, 238–247 (2002).
[23]B. Zheng, W. F. Good, D. R. Armfield, C. Cohen, T. Hertzberg, J. H. Sumkin, and D. Gur, "Performance change of mammographic CAD schemes optimized with most-recent and prior image databases," Acad. Radiol. **10**, 283–288 (2003).
[24]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," Med. Phys. **32**, 2827–2838 (2005).
[25]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, and S. Paquerault, "Evaluation of an automated computer-aided diagnosis system for the detection of masses on prior mammograms," San Diego, 2000.
[26]G. M. Te Brake, N. Karssemeijer, and J. Hendriks, "Automated detection of breast carcinomas not detected in a screening program," Radiology **207**, 465–471 (1998).
[27]D. M. Ikeda, R. L. Birdwell, K. F. O'Shaughnessy, E. A. Sickles, and R. J. Brenner, "Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography," Radiology **230**, 811–819 (2004).
[28]D. Gur, J. S. Stalder, L. A. Hardesty, B. Zheng, J. H. Sumkin, D. M. Chough, B. E. Shindel, and H. E. Rockette, "Computer-aided detection performance in mammographic examination of masses: Assessment," Radiology **233**, 418–423 (2004).
[29]H. El-Shishini, M. S. Abdel-mottaleb, M. El-Raey, and A. Shoukry, "A multistage algorithm for fast classification of patterns," Pattern Recogn. Lett. **10**, 211–215 (1989).
[30]J. Y. Zhou and T. Pavlidis, "Discrimination of characters by a multi-stage recognition process," Pattern Recogn. **27**, 1539–1549 (1994).
[31]L. K. Hansen and P. Salamon, "Neural network ensembles," IEEE Trans. Pattern Anal. Mach. Intell. **12**, 993–1001 (1990).
[32]G. Rogova, "Combining the results of several neural network classifiers," Neural Networks **7**, 777–781 (1994).
[33]S. Hashem and B. Schmeiser, "Improving model accuracy using optimal linear combinations of trained neural networks," IEEE Trans Neural Networks **6**, 792–794 (1995).
[34]T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," IEEE Trans. Pattern Anal. Mach. Intell. **16**, 66–75 (1994).
[35]D. A. Bell, J. W. Guan, and Y. Bi, "On combining classifier mass functions for text categorization," IEEE Trans. Knowl. Data Eng. **17**, 1307–1319 (2005).
[36]J. Cao, M. Ahmadi, and M. Shridhar, "Recognition of handwritten numerals with multiple feature and multistage classifier," Pattern Recogn. **28**, 153–160 (1995).
[37]A. Al-Ani and M. Deriche, "A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence," J. Artif. Intell. Res. **17**, 333–361 (2002).
[38]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, M. A. Helvie, and M. A. Roubidoux, "A dual computer-aided detection (CAD) system for improvement of mass detection on mammograms," RSNA Program Book 2004, p. 491.
[39]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, M. A. Helvie, M. A. Roubidoux, N. Petrick, C. Zhou, and J. Ge, "Computer aided detection of breast masses on mammograms: Performance improvement using a dual system," Proc. SPIE **5747**, 9–15 (2005).
[40]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, C. Zhou, and Z. Ge, "Computer aided detection of breast masses on full-field digital mammograms: False positive reduction using gradient field analysis," Proc. SPIE **5370**, 992–998 (2004).
[41]N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," Med. Phys. **26**, 1642–1654 (1999).
[42]D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," Med. Phys. **24**, 903–914 (1997).
[43]M. M. Galloway, "Texture classification using gray level run lengths," Comput. Graph. Image Process. **4**, 172–179 (1975).
[44]M. J. Norusis, *SPSS for Windows Release 6 Professional Statistics* (SPSS, Chicago, IL, 1993).
[45]L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. N. Gurcan, "Analysis of temporal change of mammographic features: Computer-aided classification of malignant and benign breast masses," Med. Phys. **28**, 2309–2317 (2001).
[46]C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat. Med. **17**, 1033–1053 (1998).
[47]B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. M. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size.," Med. Phys. **27**, 1509–1522 (2000).
[48]C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon, Oxford, 1995).
[49]J. A. Freeman and D. M. Skapura, *Neural Networks-Algorithms, Applications, and Programming Techniques* (Addison-Wesley, Reading, 1991).
[50]D. P. Chakraborty and L. H. L. Winter, "Free-response methodology: Alternate analysis and a new observer-performance experiment," Radiology **174**, 873–881 (1990).
[51]D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," Med. Phys. **31**, 2313–2330 (2004).
[52]K. F. O'Shaughnessy, R. A. Castellino, S. L. Muller, and K. Benali, "Computer-aided detection (CAD) on 90 biopsy-proven breast cancer cases acquired on a full-field digital mammography (FFDM) system," Radiology **221**(P), 471 (2001).
[53]L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology **215**, 554–562 (2000).
[54]R. E. Brem, J. W. Hoffmeister, J. A. Rapelyea, G. Zisman, K. Mohtashemi, G. Jindal, M. P. DiSimio, and S. K. Rogers, "Impact of breast density on computer-aided detection for breast cancer," AJR, Am. J. Roentgenol. **184**, 439–444 (2005).

Computer-Aided Detection Systems for Breast Masses: Comparison of Performances on Full-Field Digital Mammograms and Digitized Screen-Film Mammograms[1]

Jun Wei, PhD, Lubomir M. Hadjiiski, PhD, Berkman Sahiner, PhD, Heang-Ping Chan, PhD, Jun Ge, PhD
Marilyn A. Roubidoux, MD, Mark A. Helvie, MD, Chuan Zhou, PhD, Yi-Ta Wu, PhD
Chintana Paramagul, MD, Yiheng Zhang, PhD

Rationale and Objectives. To compare the performance of computer aided detection (CAD) systems on pairs of full-field digital mammogram (FFDM) and screen-film mammogram (SFM) obtained from the same patients.

Materials and Methods. Our CAD systems on both modalities have similar architectures that consist of five steps. For FFDMs, the input raw image is first log-transformed and enhanced by a multiresolution preprocessing scheme. For digitized SFMs, the input image is smoothed and subsampled to a pixel size of 100 μm $\times$ 100 μm. For both CAD systems, the mammogram after preprocessing undergoes a gradient field analysis followed by clustering-based region growing to identify suspicious breast structures. Each of these structures is refined in a local segmentation process. Morphologic and texture features are then extracted from each detected structure, and trained rule-based and linear discriminant analysis classifiers are used to differentiate masses from normal tissues. Two datasets, one with masses and the other without masses, were collected. The mass dataset contained 131 cases with 131 biopsy proven masses, of which 27 were malignant and 104 benign. The true locations of the masses were identified by an experienced Mammography Quality Standards Act (MQSA) radiologist. The no-mass data set contained 98 cases. The time interval between the FFDM and the corresponding SFM was 0 to 118 days.

Results. Our CAD system achieved case-based sensitivities of 70%, 80%, and 90% at 0.9, 1.5, and 2.6 false positive (FP) marks/image, respectively, on FFDMs, and the same sensitivities at 1.0, 1.4, and 2.6 FP marks/image, respectively, on SFMs.

Conclusions. The difference in the performances of our FFDM and SFM CAD systems did not achieve statistical significance.

Key Words. Computer-aided detection; mass detection; full-field digital mammogram (FFDM); screen-film mammogram (SFM); free-response receiver operating characteristic (FROC).

Acad Radiol 2007; 14:659–669

[1] From the Department of Radiology, University of Michigan, CGC B2103, 1500 E. Medical Center Drive, Ann Arbor, MI 48109 (J.W., L.M.H., B.S., H.-P.C., J.G., M.A.R., M.A.H., C.Z., Y.-T.W., C.P., Y.Z.). Received December 15, 2006; accepted February 13, 2007. This work is supported by US Army Medical Research and Materiel Command grants W81XWH-1-04-1-0475 and DAMD 17-02-1-0214, and USPHS grant CA95153. The content of this article does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, PhD, for the LABROC program and to Dev Chakraborty, PhD, for the JAFROC program. **Address correspondence to:** J.W. e-mail: jvwei@umich.edu.

doi:10.1016/j.acra.2007.02.017

Full-field digital mammography (FFDM) and screen-film mammography (SFM) are two available methods for breast cancer screening in clinical practice. FFDM detectors provide higher detective quantum efficiency (DQE) and signal-to-noise ratio (SNR), wider dynamic range, and higher contrast sensitivity than SFM. FFDM may alleviate some of the limitations of SFM, especially in breasts with dense fibroglandular tissue (1). In the last few years, several FFDM systems became commercially available because of the potential of digital imaging to improve breast cancer detection.

Several clinical trials have been conducted to compare radiologists' interpretation on FFDMs and SFMs. Lewin et al (2,3) conducted a clinical study to compare FFDMs and SFMs for the detection of breast cancer in 6,737 examinations of women 40 years of age and older collected from two institutions. Forty-two cancers were detected within this population. The difference in cancer detection was not statistically significant ($P > .1$) between FFDMs and SFMs. FFDMs resulted in fewer recalls than did SFM, which was statistically significant ($P < .001$). Another clinical trial (4) aiming at collecting data for US Food and Drug Administration approval included SFMs and FFDMs of 676 women who were scheduled to undergo breast biopsy. The average area under the receiver operating characteristic (ROC) curve, the sensitivity and the specificity were 0.715, 0.66 and 0.67 for printed FFDM and 0.765, 0.74, 0.60 for SFM, respectively. However, none of these differences achieved statistical significance. Skaane et al (5–7) has conducted several clinical studies to compare SFM and FFDM with soft-copy interpretation for reader performance in detection and classification of breast lesions. According to their findings, there was no significant difference between FFDM and SFM either in detection or in classification. A recent study by Pisano et al (1) collected a total of 49,528 patients at 33 sites in the United States and Canada. Mammograms were interpreted independently by two radiologists. The overall diagnostic accuracy of FFDMs and SFMs for breast cancers was similar. However, FFDM was more accurate in women younger than age 50 years, women with radiographically dense breasts, and premenopausal or perimenopausal women.

Studies indicate that radiologists do not detect all carcinomas that are visible on retrospective analyses of the images (8–14). Computer-aided diagnosis (CAD) is considered to be one of the promising approaches that may improve the sensitivity of mammography (15,16). Most of the mammographic CAD systems developed so far are based on digitized SFMs. Li et al (17) attempted to adapt their CAD system developed on SFMs for detection of masses on FFDMs by standardizing the FFDMs. Their preliminary results on a small data set (training on 36 normal and 24 mass cases, testing on 24 normal and 10 mass cases) showed 60% sensitivity at 2.47 false positives (FPs)/image. Several commercial CAD systems reported comparable performance on FFDMs and SFMs. However, their study was not reported in peer-reviewed journals, so that the dataset and algorithm are unknown. So far, there are no studies on comparison of breast mass detection between FFDMs and SFMs from the same patients by using CAD system. We have developed a CAD system for mass detection on SFMs (18,19) and are adapting the system to FFDMs. Our preliminary study with 65 patients was reported previously (20). In this study, we compared the performance of the two CAD systems on case-matched pairs of FFDMs and SFMs.

MATERIALS AND METHODS

Materials

Our study group consisted of patients with breast lesions that were categorized suspicious and recommended for biopsy. The patients had either FFDM or SFM for their clinical exams. Institutional review board approval and patient informed consent were obtained to acquire corresponding mammograms of the breast to be biopsied using the other modality. Therefore, the corresponding FFDM and SFM were available only from one breast for each patient. The time interval between the SFM and the FFDM ranged from 0 to 118 days. The dataset consisted of 229 patients aged 30–86 with a mean age of 55 ± 11 years. All cases have two mammographic views, the craniocaudal view and the mediolateral oblique view or the lateral view, yielding a total of 458 FFDMs and 458 corresponding SFMs. The SFMs were acquired with MinR2000 screen-film systems (Eastman Kodak, Rochester, NY) and digitized with a LUMISCAN 85 laser film scanner (Lumisys, Los Altos, CA) at a pixel resolution of 50 μm × 50 μm and 4096 gray levels. The digitizer was calibrated so that gray-level values were linearly proportional to the optical density in the range of 0–4, with a slope of 0.001 per pixel value. The digitizer output was linearly converted so that a large pixel value corresponded to a low optical density. FFDMs were acquired with a GE Senographe 2000D system (GE Medical Systems, Milwaukee, WI). The GE system has a CsI

Table 1
Description of Cases in the Mass Datasets and Subsets for Training and Testing in the Twofold Cross-Validation Scheme

	Mass Set		Mass Subset 1		Mass Subset 2	
	FFDM	SFM	FFDM	SFM	FFDM	SFM
Total number of cases	131	131	65	65	66	66
Total number of images	262	262	130	130	132	132
Number of visible masses (by case)	131	130	65	65	66	65
Number of masses only visible on one view	8	9	5	5	3	4
Number of visible masses (by image)	254	251	125	125	129	126
Number of visible malignant masses	27	27	12	12	15	15
Number of visible benign masses	104	103	53	53	51	50

FFDM: full-field digital mammogram; SFM: screen-film mammogram.

phosphor/a:Si active matrix flat panel digital detector with a pixel size of 100 μm $\times$ 100 μm and 14 bits per pixel. The raw FFDMs were used as the input of our CAD system.

The dataset included 131 cases containing masses and 98 cases containing microcalcifications without a visible mass, as determined with visual inspection by an experienced radiologist. The 131 cases will be referred to as the mass dataset and the 98 cases as the "no-mass" data set in the following discussion. The no-mass cases were considered as "normal" with respect to masses and were used to estimate the FP mark rates of the CAD systems during testing. The mass dataset contained 131 biopsy proved masses, of which 27 were malignant and 104 benign. By examining all available information, including the diagnostic mammograms and reports, the true locations of the masses were identified by an experienced Mammography Quality Standards Act (MQSA) radiologist. In these 131 mass cases, 1 mass can be seen only on FFDMs, 7 masses can be seen on only one view on both FFDMs and SFMs, and 3 masses can be seen on only one view on either FFDMs (1 mass) or SFMs (2 masses). There were therefore 131 visible masses on FFDMs and 130 visible masses on SFMs if the masses were counted by case. There were 254 visible and 8 invisible masses on FFDMs and 251 visible and 11 invisible masses on SFMs if the masses were counted independently by mammographic view. The number of images and masses in the mass dataset are described in Table 1. Figure 1 shows an example with a 7-mm malignant mass. The size of a mass was estimated as its longest diameter seen on the mammograms. The visibility of the masses was rated by the experienced radiologist on a 10-point scale, with 1 representing the most visible masses and 10 the most difficult case relative to the cases seen in clinical practice. Figures 2 and 3 show the histograms of mass sizes and visibility, respectively, for the mass set. The mass size ranged from 3 to 30 mm (mean: 12.5 $\pm$ 4.9 mm on FFDMs and 12.6 $\pm$ 4.9 mm on SFMs) and the visibility ratings extended over the entire range. Figure 4 shows the breast density in terms of BI-RADS category as estimated by the radiologist for the FFDM and SFM datasets.

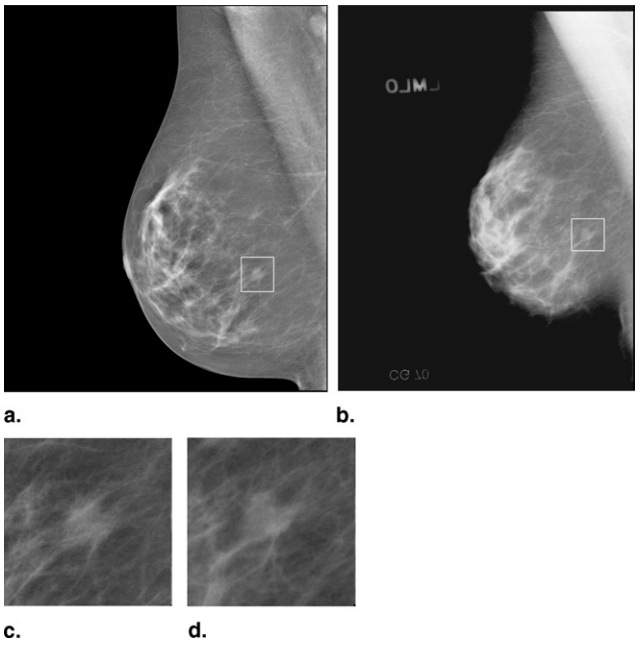


Figure 1. An example of mammograms with a region of interest (ROI) containing a malignant mass with a size of 7 mm. **(a)** Processed full-field digital mammogram (FFDM) by using the Laplacian pyramid multiscale method, **(b)** digitized screen-film mammogram (SFM), **(c)** magnified ROI on FFDM, and **(d)** magnified ROI on SFM. The SFM is displayed with the same resolution as that of the FFDM. The apparently smaller breast size on SFM is mainly caused by the very dark breast periphery region on the SFM that cannot be seen on the printed page.

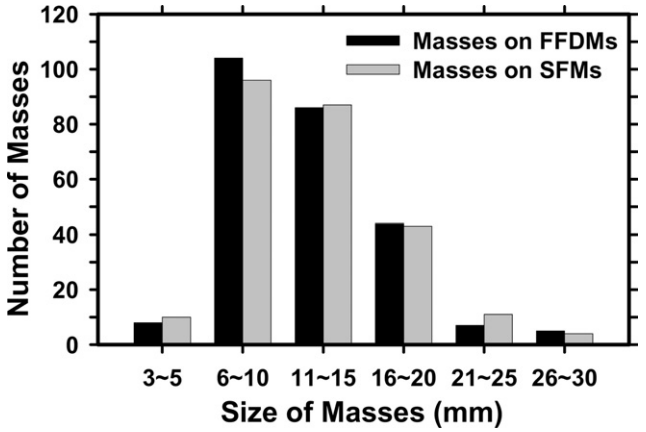


Figure 2. Histogram of the sizes for 254 masses on full-field digital mammograms (FFDMs) and 251 masses on the screen-film mammograms (SFMs) in our dataset. Mass sizes are measured as the longest dimension of the mass by an experienced Mammography Quality Standards Act (MQSA) radiologist. The size of the masses in the dataset ranged from 3 to 30 mm (mean: 12.5 ± 4.9 mm on FFDMs and 12.6 ± 4.9 mm on SFMs).

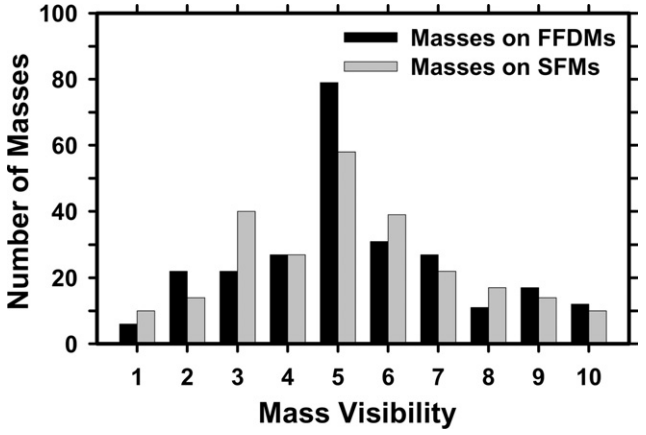


Figure 3. Histogram of the visibility of the 254 masses seen on full-field digital mammograms and 251 masses seen on screen-film mammograms in our dataset. The visibility is evaluated on a 10-point rating scale, with 1 representing the most visible masses and 10 the most difficult case relative the cases seen in their clinical practice. Each mass on a mammogram is rated independently by an experienced MQSA radiologist.

METHODS

CAD System

The major steps in the mass detection systems on FFDMs and SFMs are similar, but the feature spaces and classifiers for FP reduction in each system were designed separately to suit the characteristics of FFDMs and SFMs. The two systems are therefore described together, but the differences will be pointed out whenever applicable. Each single CAD system consists of five processing steps:

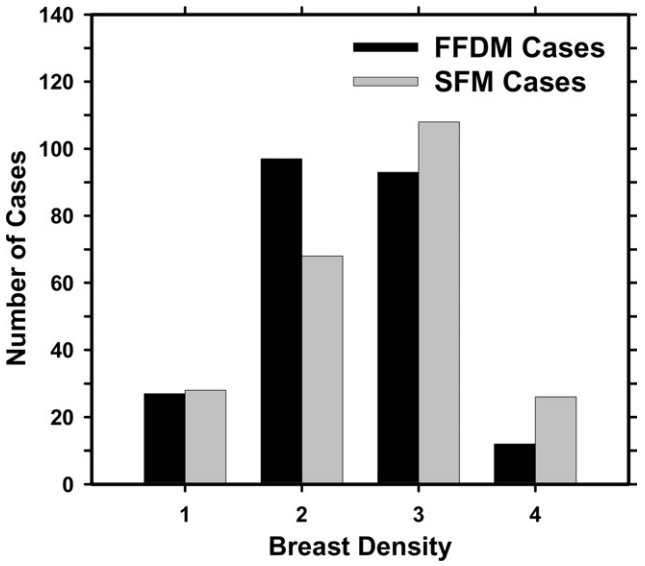


Figure 4. Distribution of the breast density for the 229 cases in terms of BI-RADS category estimated by an MQSA radiologist.

1) preprocessing, 2) prescreening of mass candidates, 3) segmentation of suspicious objects, 4) feature extraction and analysis, and 5) FP reduction by classification of normal tissue structures and masses.

FFDMs are generally preprocessed with proprietary methods by the manufacturer of the FFDM system before being displayed to readers. The image preprocessing method used depends on the manufacturer of the FFDM system. To develop a CAD system that is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we use the raw FFDM as input to our CAD system. We have previously developed a multiscale preprocessing scheme for image enhancement (21). In brief, the raw mammogram is first segmented automatically into the background and the breast region. A logarithmic transform is applied to the image which is then scaled to 12-bit. The Laplacian pyramid method (21,22) is used to decompose the transformed breast image into multiscales. A nonlinear weight function based on the pixel gray level from each of the low-pass components is designed to enhance the high-pass components. The processed image is reconstructed by summing the weighted components.

For SFMs, the full resolution digitized mammograms are smoothed with a 2×2 box filter and subsampled by a factor of 2, resulting in images having a pixel size of $100\ \mu m \times 100\ \mu m$. These images are used as input to the CAD system.

After preprocessing, a two-stage gradient field analysis method (21,23) is used to identify the mass candidates for

either FFDMs or SFMs. In brief, a gradient field analysis is employed in the first stage to identify potential mass candidates based on high values of the initial gradient field. Each potential mass candidate is segmented by a region growing technique. The shape and the gray-level information of the segmented object allow adaptive refinement of the gradient field analysis in the second stage. Locations of high radial gradient convergence are then labeled as mass candidates. These suspicious objects are segmented with a k-means clustering method (24). First, a 256×256 pixel region of interest (ROI) centered at the high gradient point is background-corrected (25) and weighted by a Gaussian function with $\sigma = 256$ pixels. K-means clustering using the pixel values in a background-corrected image and a Sobel filtered image as features is then used to segment the object.

For each suspicious object, eleven morphological features (18) are extracted. A rule-based classifier removes the detected structures that are substantially different from breast masses. Global and local multiresolution texture analyses (26) are performed in each ROI by using the spatial gray-level dependence (SGLD) matrices. Thirteen SGLD texture measures are used. Global texture features are extracted from the entire ROI for two scales, seven distances, and two angles. Local texture features are extracted from the local region containing the detected object and the peripheral regions within each ROI for two scales, four distances, and two angles. Therefore, a total of 364 features and 208 features, respectively, are extracted from global and local texture analysis. The feature space for final classification is the combination of morphologic features and SGLD texture features. Finally, linear discriminant analysis (LDA) is used to classify masses from normal tissue in the feature space. The discriminant scores are ranked for each mammogram, and any object with a discriminant score that ranks lower than three is eliminated.

Training and Test CAD System

Twofold cross-validation was used for training and testing our CAD system for FFDMs. We randomly separated the mass datasets by case into two independent subsets: subset 1 with 65 cases and subset 2 with 66 cases. The numbers of masses by image and by case for the FFDM and SFM subsets are shown in Table 1. The training included selection of proper parameters and features for the classifier in the CAD system. After the training with one mass subset was completed, the parameters and features were fixed for testing with the other mass subset. The training and test mass subsets were switched and the training and test processes were repeated. The trained CAD systems were also applied to the no-mass data set, which was not used during training, to estimate the FP rate in screening mammograms.

During training, feature selection with stepwise LDA was applied to obtain the best feature subset and reduce the dimensionality of the feature space to design an effective classifier. The detailed procedure has been described elsewhere (21,27,28). Briefly, at each step one feature was entered or removed from the feature pool by analyzing its effect on the selection criterion, which was chosen to be the Wilks' lambda in this study. Because the appropriate threshold values for feature entry, feature elimination, and tolerance of feature correlation were unknown, we used an automated simplex optimization method to search for the best combination of thresholds in the parameter space. The simplex algorithm used a leave-one-case-out resampling method within the training subset to select features and estimate the weights for the LDA classifier. To have a figure of merit to guide feature selection, the test discriminant scores from the left-out cases were analyzed using ROC methodology (29). The accuracy for classification of masses and FPs was evaluated as the area under the ROC curve, A_z, for the test cases. In this approach, feature selection was performed without the left-out case so that the test performance would be less optimistically biased (30). However, the selected feature set in each leave-one-case-out cycle could be slightly different because every cycle had one training case different from the other cycles. To obtain a single trained classifier to apply to the cross-validation test subset, a final stepwise feature selection was performed with the best combination of thresholds, found in the simplex optimization procedure, on the entire training subset to obtain the final set of features and estimate the weights of the LDA. Note that the entire process of feature selection and classifier weight estimation was performed within the training subset. The LDA classifier with the selected feature set was then fixed and applied to the cross-validation test subset. The training and testing processes were performed independently for the twofold cross-validation sets.

Because we already trained our CAD system for SFMs with a large dataset in a previous study (19), we used the trained system without retraining the parameters in this study. For testing, we divided the SFMs into two test datasets that followed the same case grouping as that for FFDMs. The test cases in each subset did not overlap

with any training cases used for training the SFM CAD system in the previous study.

Evaluation Methods

We used a free-response ROC (FROC) method (31) to assess the overall performance of the CAD scheme on this image set. A FROC curve is obtained by plotting the mass detection sensitivity as a function of FP marks per image as the decision threshold on the LDA classifier scores varies.

The detected individual objects were compared with the "true" mass locations marked by the experienced radiologist, as described previously. A detected object was labeled as true positive (TP) if the overlap between the bounding box of the detected object and the bounding box of the true mass relative to the larger of the two bounding boxes was over 25%. Otherwise, it would be labeled as FP. The 25% threshold was selected as described in our previous study (18).

FROC curves were presented on a per-image and a per-case basis. For image-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to the number of masses by image on each subset of FFDMs or SFMs (Table 1). For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either or both masses on the two views was considered to be a TP detection; the sensitivity was thus calculated relative to the number of masses by case on each subset of FFDMs or SFMs (Table 1). The test FROC curve for a given mass subset was estimated by counting the detected masses on the test mass subset for the sensitivity. The FP marker rate was estimated in two ways: one from FPs detected in the same test mass subsets, the other from FPs detected in the no-mass dataset. For the latter, we applied the trained CAD system to the entire no-mass dataset. The average number of FP marks per image produced by the CAD system at a given sensitivity was estimated by counting the detected objects in these cases at the corresponding decision threshold. Because we used twofold cross-validation method for training and testing, we obtained two test FROC curves, one for each test subset, for each of the modalities. To summarize the results for comparison, an average test FROC curve was derived by averaging the FP rates at the same sensitivity along the FROC curves of the two corresponding test subsets.

To compare the performance of our CAD system for FFDMs and SFMs statistically, we applied the alternative free-response ROC (AFROC) method and the jackknife free-response ROC (JAFROC) method developed by Chakraborty et al (32,33) to the pairs of FROC curves. In the AFROC method, the FROC data are first transformed by counting the number of false-positive images instead of the FPs per image. The LDA score of a false-positive image is determined by the highest score FP object on the image regardless of how many lower scores FP objects are made on the same image. The ROCKIT curve fitting software and statistical significance tests for ROC analysis developed by Metz et al (29) can then be used to analyze the AFROC data.

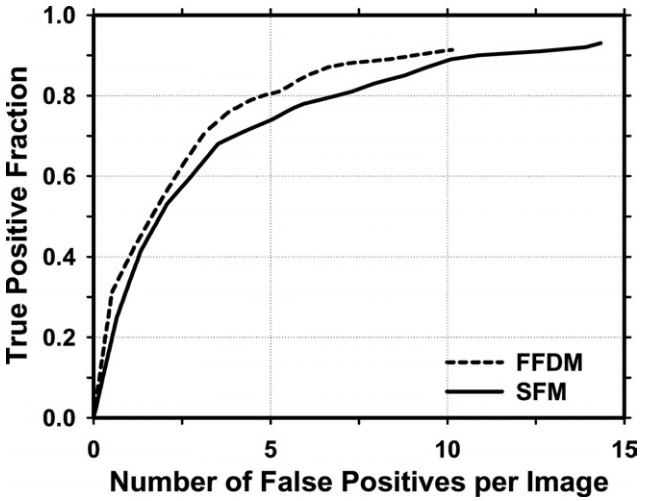


Figure 5. Comparison of free-response receiver operating characteristic (FROC) curves on full-field digital mammograms and screen-film mammograms during the prescreening stage. The FROC curves were generated by varying the number of detected suspicious objects per image based on the ranking of the local maxima on gradient field images. The FP rate was estimated from the mammograms with masses.

RESULTS

For simplicity, we combined the detection results on the two test subsets from the twofold cross-validation process in the following discussion. The prescreening stage detected 91.3% (232/254) of the masses with an average of 10.13 (2,655/262) FPs /image on FFDMs and 93.2% (234/251) with an average of 14.43 (3,781/262) FPs/image on SFMs. Figure 5 compares the FROC curves on FFDMs and SFMs during the prescreening stage. The FROC curves were generated by varying the number of detected suspicious objects per image based on the ranking of local maxima on the gradient field images.

We used two steps for FP reduction for both CAD systems. The first step was the rule-based classification based on morphologic features. After this step, there were 2,572 mass candidates (9.8 objects/image) on FFDMs and 3,654 mass candidates (13.9 objects/image) on SFMs without additional FNs for the test sets of 262 images. The second step was the LDA classification. A total of 16 (4 global texture features, 7 local texture features, and 5 morphologic features) and 12 (4 global texture features, 4 local texture features, and 4 morphologic features) features, respectively, were selected from the two independent training subsets for FFDMs. The feature set for SFMs contained a total of 21 features (11 global texture features, 7 local texture features, and 3 morphologic features), as obtained from previous training.

Figure 6 shows the comparison of the average test FROC curves of the CAD systems for FFDMs and SFMs. The FFDM CAD system achieved a case-based sensitivity of 70%, 80%, and 90% at 0.67, 1.15, and 1.93 FPs/image, respectively, compared with 0.75, 1.06, and 1.86 FPs/image for the SFM CAD system. Because two trained CAD systems were obtained for the FFDMs from the cross-validation training, we applied each of the trained systems to the no-mass data set for FROC analysis, and estimated the number of FP marks per image on the no-mass cases at each decision threshold. For each trained CAD system, the sensitivity was estimated from the detected masses on the test mass subset and plotted against the FP rate estimated from the no-mass set. Figure 7 shows the average FROC curves for FFDMs and SFMs, similar to those shown in Fig 6, except that the FP rates were estimated from the no-mass data set.

The comparison of the FROC curves for the FFDM and SFM CAD systems in terms of the area under the fitted AFROC curve (A_1) and the P values for both test mass subsets are summarized in Table 2. The differences in the A_1 values between the two modalities did not achieve statistical significance ($P > .05$). The fitted AFROC curves, however, did not fit very well to the transformed AFROC data, as discussed previously (21). For the JAFROC method, Chakraborty et al provided software to estimate the statistical significance of the difference between two FROC curves. The comparison of the figure-of-merit (FOM) and the P values is also summarized in Table 2. The differences in the FOMs between the FFDM and SFM CAD systems again did not achieve statistical significance ($P > .05$).

There were 27 malignant cases in the mass set. Figure 8 compares the average test FROC curves of the FFDM and SFM CAD systems for detection of malignant masses. The FP rate was estimated from the no-mass dataset. In this case, the FFDM CAD system achieved a case-based sensitivity of 70%, 80%, and 90% at 0.37, 0.73, and 1.31 FP marks/image, respectively, which were substantially better than the FP rates of 1.1, 1.6, and 2.0 FP marks/image for the SFM CAD system. However, the difference did not achieve statistical significance ($P > .05$).

A total of 105 FFDM cases and 134 SFM cases were identified as BI-RADS 3 and 4 categories by an MQSA

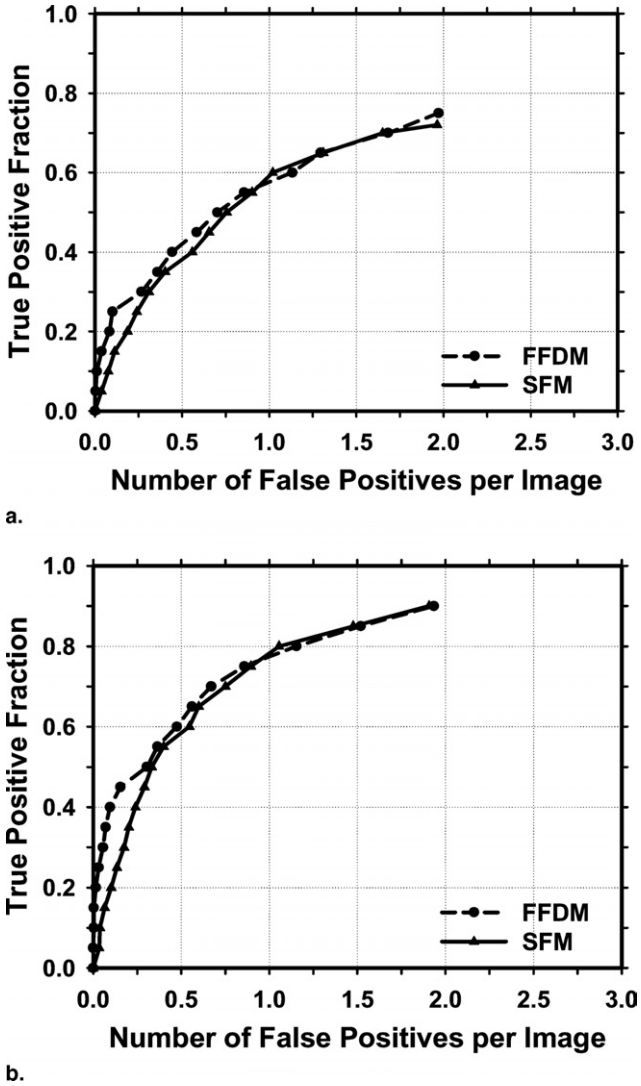
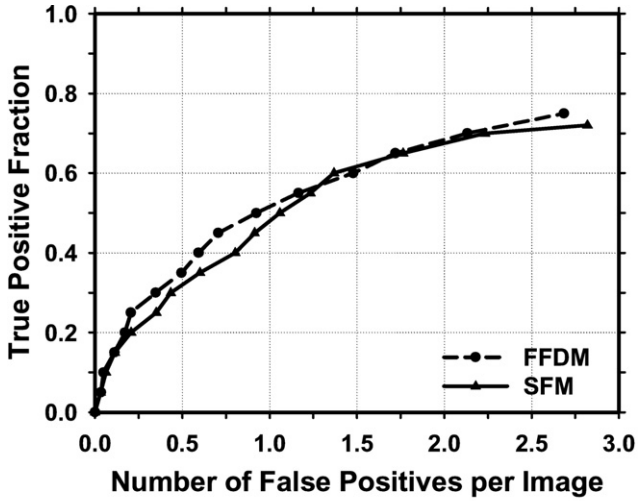
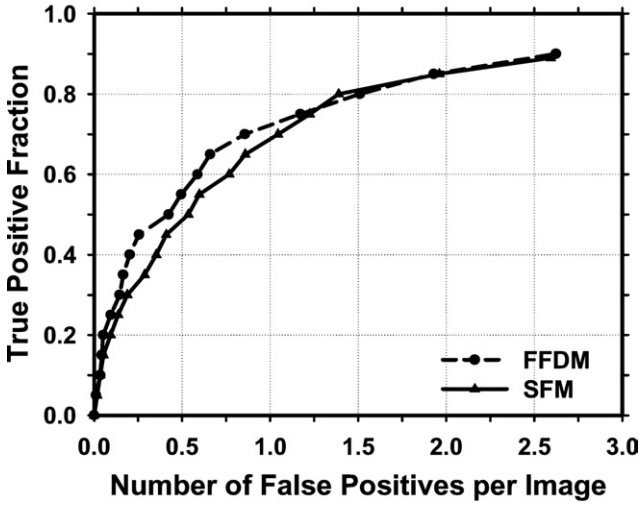


Figure 6. Comparison of the average test free-response receiver operating characteristic (FROC) curves obtained from averaging the FROC curves of the two independent mass subsets on full-field digital mammograms and screen-film mammograms. The FP rate was estimated from the mammograms with masses. **(a)** Image-based FROC curves and **(b)** case-based FROC curves.



a.



b.

Figure 7. Comparison of the average test free-response receiver operating characteristic (FROC) curves obtained from averaging the FROC curves of the two independent mass subsets on full-field digital mammograms and screen-film mammograms. The FP rate was estimated from the mammograms without masses. **(a)** Image-based FROC curves and **(b)** case-based FROC curves.

radiologist (Fig 4). Of these, 88 cases (56 mass cases and 32 no-mass cases) were in common. Figure 9 compares the average test FROC curves of the FFDM and SFM CAD systems for detection of masses only on this common subset of dense breasts. The FP rate was estimated from the 32 no-mass dense breasts. Although the FROC curve for the FFDMs appears to be slightly higher than that of the SFMs, the difference did not achieve statistical significance ($P > .05$).

DISCUSSION

CAD systems have been proven to be helpful as a second opinion to assist radiologists in interpretation of SFMs. Recently several studies have been conducted to compare FFDM with SFM in screening cohorts (1,4,5,34). These clinical trials arrived at different conclusions about the advantages or disadvantages of FFDM in comparison to conventional SFM systems. Some of the differences may be attributed to factors such as the mammographic equipment, the study design, the sample sizes, and the reader experience. It is also important to compare the performances of FFDM and SFM CAD systems. In our study, we compared the performance of the two systems on pairs of FFDM and SFM obtained from the same patients at close time intervals.

Several FFDM systems have been approved for clinical applications. Because digital detectors generally have a linear response to x-ray exposure, the raw pixel values are a linear function of the absorbed x-ray energy in the detector. To develop a CAD system that is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM as input to our CAD system. Although the spatial resolution and noise properties of the images from different detectors were still different, the use of raw images already reduced one of the major differences between mammograms from different FFDM systems. For preprocessing of the raw FFDMs, we developed a multiresolution enhancement method. From our observation on the SFMs and the processed FFDMs, the breast tissue on SFMs appears to be denser than that on FFDMs (35). This may be attributed to the harder beam quality used and the Laplacian enhancement on FFDMs. In this study, 134 SFM cases were rated as BI-RADS 3 and 4 categories by an MQSA radiologist, whereas only 105 FFDM cases were rated as BI-RADS 3 and 4. When the FFDM and SFM CAD systems were applied to the small common subset (56 with masses and 32 without masses) of dense breasts rated as BI-RADS 3 and 4, there was no significant difference between their average test FROC curves (Fig 9).

The overall performances of the CAD systems for the two modalities did not demonstrate significant difference for comparisons in either the subsets or the entire dataset. One factor may be the substantially smaller number of training samples used for the FFDM CAD system than that for the SFM CAD system, which was trained with a set of 486 SFMs in a previous study (19). We have

Table 2
Estimation of the Statistical Significance of the Difference in the FROC Performances Between the FFDM and SFM CAD Systems

	A_1 (AFROC)				FOM (JAFROC)			
	All Cases		Malignant Cases		All Cases		Malignant Cases	
	Test Subset 1	Test Subset 2	Test Subset 1	Test Subset 2	Test Subset 1	Test Subset 2	Test Subset 1	Test Subset 2
FFDM	0.48	0.49	0.51	0.49	0.47	0.48	0.55	0.47
SFM	0.42	0.43	0.47	0.42	0.46	0.41	0.48	0.42
P values	.17	.16	.56	.23	.73	.33	.29	.59

The FROC curves with the FP marker rates obtained from the no-mass dataset were compared.

FROC, ; FFDM, full-field digital mammogram; SFM, screen-film mammogram; CAD, computed-aided detection; AFROC, alternative free-response receiver operating characteristic; FOM, figure-of-merit; JAFROC, jackknife free-response ROC.

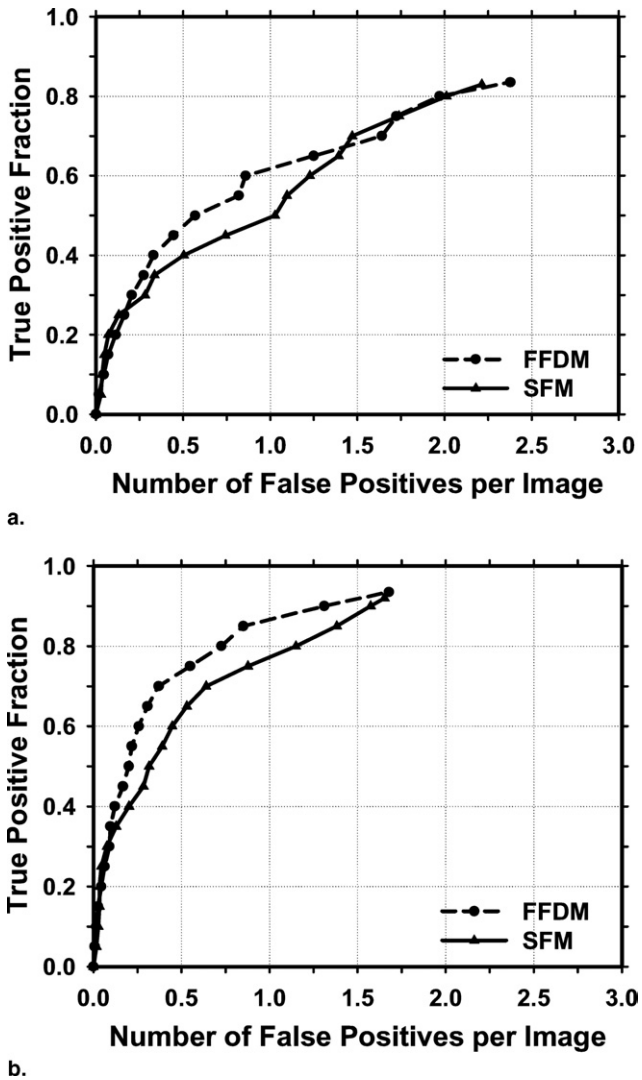


Figure 8. Comparison of the average test free-response receiver operating characteristic (FROC) curves of computed-aided detection systems on full-field digital mammograms and screen-film mammograms for mammograms with malignant masses. The FP rate was estimated from the mammograms without masses. **(a)** Image-based FROC curves and **(b)** case-based FROC curves.

shown previously that a classifier designed with a larger number of training samples will have better generalization to unknown test cases (36). Furthermore, because our CAD system was originally developed on SFMs, some of those techniques used may favor SFMs. If new techniques are designed to specifically suit the properties of FFDMs, the biases may be reduced. Further investigations are underway to improve the FFDM CAD system.

We used a twofold cross-validation method for training and testing of the CAD systems. Feature selection and classifier weight design were performed within the training subset and thus were independent of the test subset. Kupinski et al (37) showed that feature selection and classifier weight design using the same training set of a limited size will introduce additional optimistic bias to the training result and thus additional pessimistic bias to the test result. Under the constraint of a limited training set, the relative gain or loss in terms of bias if the training set is further split into two subsets for separate feature selection and classifier weight design in comparison to using the entire set of available training samples for both processes is still unknown. The relative efficiency of different resampling techniques in utilization of a limited dataset for classifier design with or without feature selection remains an important area of further studies. In screening mammography, the cancer rate is about 3–5 per 1,000. Most of the mammograms are normal. Therefore, some CAD researchers and users estimate the FP rate using normal mammograms (38–40) because it reflects how the CAD system performs in terms of specificity in a screening setting. Furthermore, for CAD systems that set a maximum number of detected objects at the output, estimating the number of FPs using images with lesions can potentially lead to an optimistic bias for the FROC curve because one of the detected objects will likely be the true

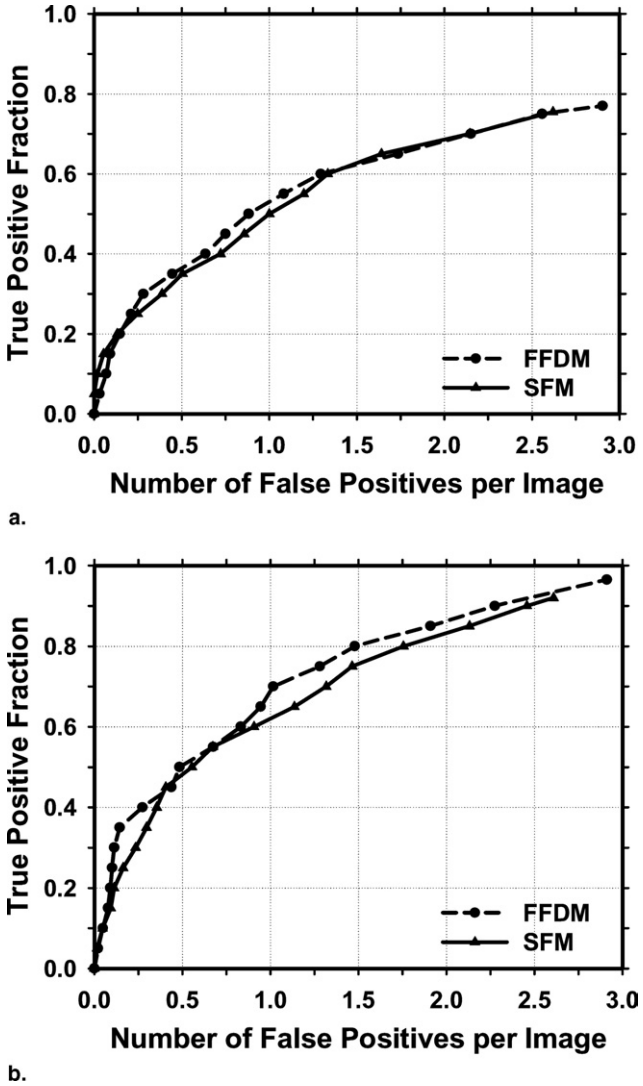


Figure 9. Comparison of the average test free-response receiver operating characteristic (FROC) curves of computed-aided detection systems on full-field digital mammograms and screen-film mammograms for the common subset of 56 dense breasts with masses rated as BI-RADS 3 and 4. The FP rate was estimated from 32 no-mass dense breasts that were also rated as BI-RADS 3 and 4. **(a)** Image-based FROC curves and **(b)** case-based FROC curves.

lesion. The FP rate can thus be underestimated by as much as 1 per image. In addition, the JAFROC analysis requires that the FP rates be estimated on normal images. We therefore reported the FP rates of our CAD systems on both mammograms with masses and without masses to facilitate comparison with other CAD systems in case investigators may evaluate their FP rates in either way.

Although we collected case-matched cases for comparing the performances of the CAD systems for FFDMs and SFMs, the images may not be exactly matched. Variations from positioning, compression force, and the difference in time between the two acquisitions would cause differences in the subtlety of the masses on the FFDMs and SFMs. However, assuming that the differences are random, both datasets would include images that have better or worse positioning, for example, than that on the other modality. The differences in the various factors would likely be averaged out over the entire dataset. We expect that they might not cause substantial bias in the comparison of the relative performances of the CAD systems for the two modalities.

For a CAD system, its performance for detecting malignant masses is more important than its performance for detecting all masses. We only have 27 malignant cases in this dataset. Although the FROC curves for detection of malignant masses (Fig 8) indicated that the FFDM CAD system had a higher sensitivity than that of the SFM CAD system, the differences in the A_1 and the FOM did not achieve statistical significance ($P > .05$) for either test subsets, as shown in Table 2. A large dataset is being collected for further comparison of the FFDM and SFM CAD systems for breast cancer cases.

Conclusion

We compared the performance of our CAD systems for detection of breast masses on case-matched FFDM images and SFM images. The two CAD systems used similar computer vision techniques but their preprocessing methods were different and the FP classifiers were separately trained to adapt to the image properties of each modality. From the comparison of FROC curves, it was found that the FFDM CAD system achieved higher detection sensitivity than the SFM CAD system at the same FP rates for malignant cases. However, the performances of our FFDM and SFM CAD systems for the entire data set were similar. The differences between the two modalities were not statistically significant with both AFROC and JAFROC methods for either the entire dataset or the malignant cases alone. Further study is under way to collect a larger dataset and to improve the performances of both systems.

REFERENCES

1. Pisano ED, Gasonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005; 353:1773–1783.
2. Lewin JM, Hendrick RE, D'Orsl CJ, et al. Comparison of full-field digital mammography with screen-film mammography for cancer detection: results of 4,945 paired examinations. Radiology 2001; 218:873–880.

3. Lewin JM, D'Orsi CJ, Hendrick RE, et al. Clinical comparison of full-field digital mammography and scree-film mammography for detection of breast cancer. AJR Am J Roentgenol 2002; 179:671–677.
4. Cole E, Pisano E, Brown M, et al. Diagnostic accuracy of Fischer SenoScan digital mammography versus screen-film mammography in a diagnostic mammography population. Acad Radiol 2004; 11:876–880.
5. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading—Oslo I Study. Radiology 2003; 229: 877–884.
6. Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program—The Oslo II Study. Radiology 2004; 232:197–204.
7. Skaane P, Balleyguier C, Diekmann F, et al. Breast lesion detection and classification: comparison of screen-film mammography and full-field digital mammography with soft-copy reading—observer performance study. Radiology 2005; 237:37–44.
8. Hillman BJ, Fajardo LL, Hunter TB, et al. Mammogram interpretation by physician assistants. AJR Am J Roentgenol 1987; 149:907–911.
9. Bassett LW, Bunnell DH, Jahanshahi R, et al. Breast cancer detection: one versus two views. Radiology 1987; 165:95–97.
10. Wallis MG, Walsh MT, Lee JR. A review of false negative mammography in a symptomatic population. Clin Radiol 1991; 44:13–15.
11. Harvey JA, Fajardo LL, Innis CA. Previous mammograms in patients with impalpable breast carcinomas: retrospective vs blinded interpretation. AJR Am J Roentgenol 1993; 161:1167–1172.
12. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992; 184:613–617.
13. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists—findings from a national sample. Arch Intern Med 1996; 156:209–213.
14. Beam V, Sullivan D, Layde P. Effect of human variability on independent double reading in screening mammography. Acad Radiol 1996; 3:891–897.
15. Shtern F, Stelling C, Goldberg B, et al. Novel technologies in breast imaging: National Cancer Institute perspective. In: Society of Breast Imaging Conference. Orlando, FL, 1995; 153–156.
16. Vyborny CJ. Can computers help radiologists read mammograms? Radiology 1994; 191:315–317.
17. Li L, Clark RA, Thomas JA. Computer-aided diagnosis of masses with full-field digital mammography. Acad Radiol 2002; 9:4–12.
18. Petrick N, Chan HP, Sahiner B, et al. Breast cancer detection: evaluation of a mass detection algorithm for computer-aided diagnosis: experience in 263 patients. Radiology 2002; 224:217–224.
19. Wei J, Sahiner B, Hadjiiski LM, et al. Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms. SPIE Proc 2006; 6144:241–247.
20. Wei J, Sahiner B, Chan HP, et al. Computer aided diagnosis system for mass detection: comparison of performance on full-field digital mammograms and digitized film mammograms. RSNA 2003, Chicago, November 30–December 5, 2003:387.
21. Wei J, Sahiner B, Hadjiiski LM, et al. Computer aided detection of breast masses on full field digital mammograms. Med Phys 2005; 32: 2827–2838.
22. Burt PJ, Adelson EH. The Laplacian pyramid as a compact image code. IEEE Trans Commun 1983; COM-31:337–345.
23. Wei J, Sahiner B, Hadjiiski LM, et al. Computer aided detection of breast masses on full-field digital mammograms: false positive reduction using gradient field analysis. Proc SPIE Med Imaging 2004; 5370: 992–998.
24. Petrick N, Chan HP, Sahiner B, et al. Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms. Med Phys 1999; 26:1642–1654.
25. Sahiner B, Petrick N, Chan HP, et al. Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization. IEEE Tran Med Imaging 2001; 20:1275–1284.
26. Wei D, Chan HP, Petrick N, et al. False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis. Med Phys 1997; 24:903–914.
27. Norusis MJ. SPSS for Windows release 6 professional statistics. Chicago, IL: SPSS Inc, 1993.
28. Hadjiiski LM, Sahiner B, Chan HP, et al. Analysis of temporal change of mammographic features: computer-aided classification of malignant and benign breast masses. Med Phys 2001; 28:2309–2317.
29. Metz CE, Herman BA, Shen JH. Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. Stat Med 1998; 17:1033–1053.
30. Sahiner B, Chan HP, Petrick N, et al. Feature selection and classifier performance in computer-aided diagnosis: the effect of finite sample size. Med Phys 2000; 27:1509–1522.
31. Swensson RG. Unified measurement of observer performance in detection and localizing target objects on images. Med Phys 1996; 23: 1709–1724.
32. Chakraborty DP, Winter LHL. Free-response methodology: alternate analysis and a new observer-performance experiment. Radiology 1990; 174:873–881.
33. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys 2004; 31:2313–2330.
34. Hendrick R, Lewin J, D'Orsi C, et al. Non-inferiority study of FFDM in an enriched diagnostic cohort: comparison with screen-film mammography in 625 women. In: Yaffe MJ, ed. IWDM 2000: 5th International Workshop on Digital Mammography: Medical Physics, 2001; 475–481.
35. Chan HP, Wei J, Zhou C, Helvie MA, et al. Comparison of mammographic density estimated on digital mammograms and screen-film mammograms. RSNA 89th Scientific Assembly (Chicago, IL) 2003:424.
36. Chan HP, Sahiner B, Wagner RF, et al. Classifier design for computer-aided diagnosis: effects of finite sample size on the mean performance of classical and neural network classifiers. Med Phys 1999; 26:2654–2668.
37. Kupinski MA, Giger ML. Feature selection with limited datasets. Med Phys 1999; 26:2176–2182.
38. O'Shaughnessy KF, Castellino RA, Muller SL, et al. Computer-aided detection (CAD) on 90 biopsy-proven breast cancer cases acquired on a full-field digital mammography (FFDM) system. Radiology 2001; 221: 471.
39. Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology 2000; 215:554–562.
40. Brem RE, Hoffmeister JW, Rapelyea JA, et al. Impact of breast density on computer-aided detection for breast cancer. AJR Am J Roentgenol 2005; 184:439–444.

Bilateral analysis based false positive reduction for computer-aided mass detection

Yi-Ta Wu,[a] Jun Wei, Lubomir M. Hadjiiski, Berkman Sahiner, Chuan Zhou, Jun Ge, Jiazheng Shi, Yiheng Zhang, and Heang-Ping Chan
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109

(Received 13 February 2007; revised 20 May 2007; accepted for publication 18 June 2007; published 26 July 2007)

We have developed a false positive (FP) reduction method based on analysis of bilateral mammograms for computerized mass detection systems. The mass candidates on each view were first detected by our unilateral computer-aided detection (CAD) system. For each detected object, a regional registration technique was used to define a region of interest (ROI) that is "symmetrical" to the object location on the contralateral mammogram. Texture features derived from the spatial gray level dependence matrices and morphological features were extracted from the ROI containing the detected object on a mammogram and its corresponding ROI on the contralateral mammogram. Bilateral features were then generated from corresponding pairs of unilateral features for each object. Two linear discriminant analysis (LDA) classifiers were trained from the unilateral and the bilateral feature spaces, respectively. Finally, the scores from the unilateral LDA classifier and the bilateral LDA asymmetry classifier were fused with a third LDA whose output score was used to distinguish true mass from FPs. A data set of 341 cases of bilateral two-view mammograms was used in this study, of which 276 cases with 552 bilateral pairs contained 110 malignant and 166 benign biopsy-proven masses and 65 cases with 130 bilateral pairs were normal. The mass data set was divided into two subsets for twofold cross-validation training and testing. The normal data set was used for estimation of FP rates. It was found that our bilateral CAD system achieved a case-based sensitivity of 70%, 80%, and 85% at average FP rates of 0.35, 0.75, and 0.95 FPs/image, respectively, on the test data sets with malignant masses. In comparison to the average FP rates for the unilateral CAD system of 0.58, 1.33, and 1.63, respectively, at the corresponding sensitivities, the FP rates were reduced by 40%, 44%, and 42% with the bilateral symmetry information. The improvement was statistically significance ($p<0.05$) as estimated by JAFROC analysis. [DOI: 10.1118/1.2756612]

Key words: computer-aided detection (CAD), bilateral analysis, mass detection, false positive reduction

I. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 to 55 years of age.[1] It has been reported that early diagnosis and treatment can significantly improve the chance of survival for patients with breast cancer.[2–4] Although mammography is a powerful screening tool for detecting breast cancer,[5,6] studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially.[7–9] It has been shown that computer-aided detection (CAD) can increase the cancer detection rate by radiologists both in the laboratory and in clinical practice.[10–15]

In screening mammography, two mammographic views, cranio-caudal (CC) and mediolateral oblique (MLO) views are generally taken of each breast. During mammographic interpretation, the radiologist combines complex information including morphology, texture, and geometric location of any suspicious structures of the imaged breast from different views, asymmetric density patterns between bilateral mammograms of the same view, and changes between the current and the prior mammograms if available. Radiologists have found that these techniques are effective in improving the accuracy of detecting subtle lesions and reducing false positives (FPs).

Investigators have attempted to implement the multiple image techniques in CAD systems to improve the detection accuracy of abnormalities and the classification accuracy of differentiating malignant and benign lesions. Hadjiiski *et al.*[16] developed an interval change analysis of masses on current and prior mammograms and found that the classification accuracy of masses can be improved significantly in comparison to single image classification. Paquerault *et al.*[17] developed a two-view (CC and MLO views) fusion technique to reduce FPs in mass detection and obtained significant improvement by comparing to their one-view detection system. van Engeland *et al.*[18] recently presented a two-view CAD system by using the features including the difference in the radial distance from the candidate regions to the nipple, the gray scale correlation between both regions, and the mass likelihood of the regions determined by the single view CAD scheme. Yin *et al.*[19] used bilateral subtraction in a prescreening step of a mass detection program to locate mass candidates, but the subsequent image analysis was performed

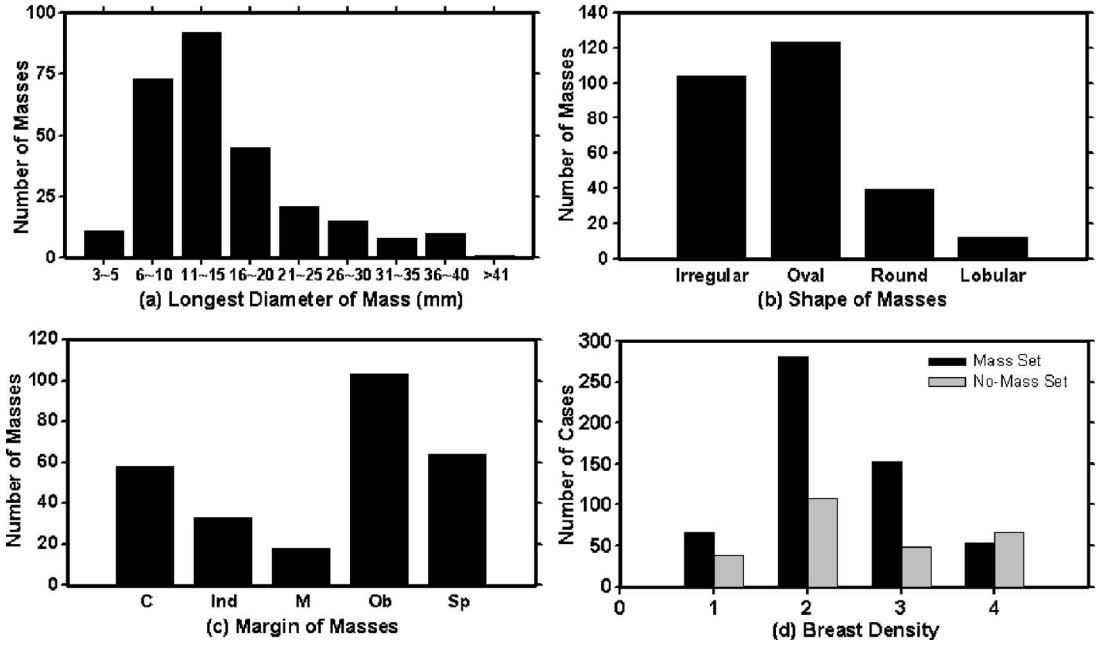


FIG. 1. The characteristics of our mass data set: (a) distribution of mass sizes, (b) distribution of mass shapes, (c) distribution of mass margins, *C*: circumscribed, Ind: indistinct, *M*: microlobulated, Ob: obscured, Sp: spiculated, and (d) distribution of the breast density in terms of BI-RADS category estimated by a MQSA radiologist.

based only on a single view. Mendez *et al.*[20] developed a bilateral CAD system based on a bilateral subtraction approach and used size and eccentricity tests and texture features to eliminate FPs. Again, the bilateral information is only used to find the suspicious objects and the subsequent analysis is based on a single view.

The detection of masses on mammograms is a challenging task. The normal fibroglandular tissue in the breast causes FPs by mimicking masses and causes false negatives (FNs) due to overlapping with lesions. In order to improve the performance of our mass detection system, we are investigating computer-vision methods by incorporating information from two-view mammograms[17] and bilateral mammograms,[21] emulating radiologists' mammographic interpretation techniques. In this study, we will discuss our approach to FP reduction by analyzing the symmetry or asymmetry of density patterns between bilateral mammograms.

II. MATERIALS AND METHODS

A. Data sets

A database of mammograms was collected from patient files at the Department of Radiology with Institutional Review Board approval. The mammograms were digitized by a Lumiscan laser scanner with a pixel size of 50 μm $\times$50 μm and 12 bits per pixel. The pixel size was increased to 100 μm$\times$100 μm by averaging every 2$\times$2 adjacent pixels before being input to the CAD system. In this study, two data sets are used: a mass data set containing bilateral digitized mammograms with malignant or benign masses and a no-mass data set containing bilateral digitized mammograms without masses, verified by an experienced radiologist. All cases had four mammographic views, the CC view and the MLO view mammogram for both breasts. The mass set and the no-mass data set contained 276 cases (552 bilateral pairs) and 65 cases (130 bilateral pairs), respectively, yielding a total of 1364 mammograms. The mass data set was used to estimate the detection sensitivity and the no-mass data set was used for estimating the FP rate (number of FPs per image). In the mass data set, each patient had a biopsy-proven mass in one of the breasts, resulting in a total of 276 masses, 166 of which were benign and 110 malignant. A Mammography Quality Standard Act (MQSA) radiologist identified the location of the masses based on all available diagnostic and clinical information of the case, measured the mass sizes as the longest dimension seen on the two-view mammograms, provided descriptors of the mass shapes and mass margins, and also provided an estimate of the breast density in term of Breast Imaging Reporting and Database System (BI-RADS) category. Figure 1 shows the information of our data set which includes the distributions of mass sizes, mass shapes, mass margins, and breast density.

For training and evaluation of the performances of the CAD systems, the cases in our mass data set were divided into two independent data subsets containing 136 and 140 cases, respectively, for twofold cross-validation training and testing. Of the 136 cases in subset 1, 52 were malignant and 84 were benign. Of the 140 cases in subset 2, 58 were malignant and 82 were benign. The no-mass data set was not used during training. All 260 mammograms were kept as independent test samples to be used with both test subsets.

B. Methods

Our bilateral CAD system combines unilateral features with bilateral features to reduce FPs. Similar structures that

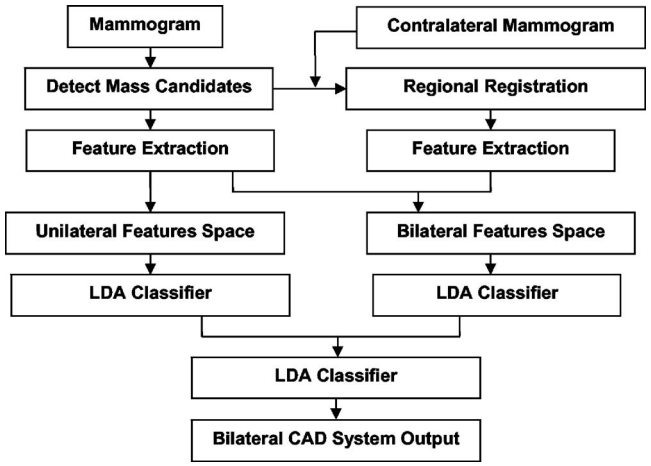


FIG. 2. Block diagram of the bilateral CAD system for mass detection on mammograms.

appear in both right and left mammograms at corresponding locations are more likely to be normal tissue than masses, whereas asymmetric density may indicate a developing lesion. The key of this system is therefore the design of a classifier that can differentiate symmetry and asymmetry of paired regions of interest (ROIs) in corresponding regions on bilateral mammograms of the same view. The system consists of four steps: (1) mass candidate (MC) localization, (2) corresponding ROIs (CR) registration, (3) feature extraction and analysis, and (4) bilateral information fusion. Figure 2 shows the block diagram for our bilateral CAD system. The detailed description for each step is presented below.

1. Mass candidate localization

Identification of mass candidates is performed by the following two steps: breast segmentation and mass candidate detection. The breast image is first segmented from the surrounding image background by boundary detection.

The algorithm developed by Zhou *et al.*[22] in our laboratory is used to track the breast boundary and segment the breast from the background. Mass detection is performed only in the breast region. We have previously developed a mass detection system for unilateral mammograms.[23–25] The system is used for mass candidate detection in the current study. The system performs mass detection in two steps. In the first step, a gradient field analysis method is used to determine the seeds of mass candidates followed by a region growing[24] method to segment the mass candidates starting from those seeds. In the second step, the gradient convergence is calculated using the gray levels and the shape of the segmented mass region as *a priori* information. The mass candidates that pass the gradient convergence criterion are retained for further analysis in the bilateral system. Figure 3 shows an example of mass candidates detected on a mammogram. Figures 3(a)–3(c) show the original image, detected breast boundary, and the detected mass candidates, respectively.

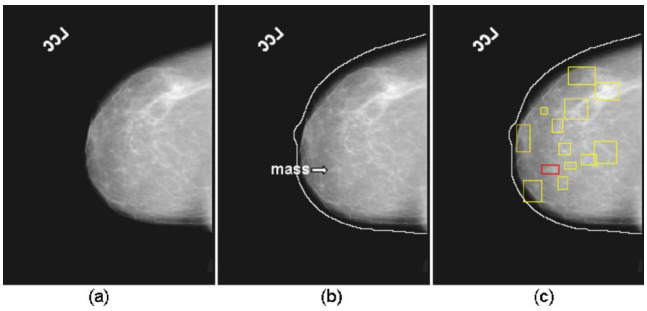


FIG. 3. An example of performing the mass candidate identification: (a) an original mammogram, (b) the detected breast boundary of (a), a mass is marked by the arrow, and (c) the detected mass candidates of (a).

2. Corresponding ROI registration

For each mass candidate, its corresponding ROI on the contralateral mammogram is identified by the regional registration technique developed previously in our laboratory[16] with a modification to handle the special case when the distance between the nipple location and the center of a ROI is too small to obtain the intersection points on the breast boundary. The nipple location on each image was manually identified so that the effectiveness of the bilateral analysis method could be evaluated independent of nipple detection errors.

The original region registration technique included the following steps. The registration is performed in a polar coordinate system where the origin is located at the nipple location of a breast image. Figure 4 shows an example of locating the corresponding ROI of a mass candidate on the contralateral mammogram. Using the distance r from the nipple o to the center of the mass as the radius, an arc centered at the origin (nipple) is drawn. The arc will intersect the mass candidate and the breast boundary at two points, p and q. The angle between **om** and **op** is defined as θ, the angle between **op** and **oq** is defined as α. On the contralateral mammogram, the corresponding ROI m' is localized with a similar procedure. An arc of radius r centered at the nipple o' of the contralateral mammogram is drawn. The intersections of the arc with the breast boundary are p' and q'. The angle

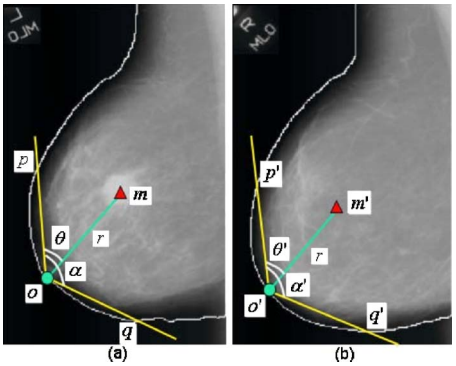


FIG. 4. An example of obtaining the corresponding ROI of a mass candidate on the contralateral mammogram: (a) mass candidate on the left MLO view at m and (b) corresponding ROI on the right MLO view at m'.

Table of Contents

merged by a third LDA. The weights of this LDA classifier were also trained with the training subset. The output score from the third LDA is used to differentiate true positives (TPs) from FPs in the bilateral CAD system.

5. Evaluation methods

The detected individual objects were compared with the true mass location marked by an experienced radiologist. An object was considered to be a TP if the overlap between the detected object and the true mass was greater than 25%. The 25% threshold was selected as described in our previous study.[30]

To evaluate the performance of our bilateral LDA classifier, the test discriminant scores were analyzed using receiver operating characteristic (ROC) methodology.[31] The accuracy for classification of mass and normal tissue was evaluated as the area under the ROC curve, A_z.

The detection performance of the bilateral CAD system was assessed by free response ROC (FROC) analysis. A FROC curve shows the relationship between the detection sensitivity and the FP rate as the decision threshold varies. FROC curves were presented on a per-image and a per-case basis. For image-based FROC analysis, the mass on each mammogram was considered an independent true object. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of the masses on either view or on both views was considered to be a TP detection.

Two sets of trained parameters were acquired as a result of the twofold cross-validation training. To estimate the FP rate on normal mammograms when the trained CAD system is used in a screening setting, we applied the trained unilateral and bilateral systems to the 260 no-mass mammograms for independent testing. The number of FP marks produced by the algorithm was estimated by counting the detected objects on these normal cases only. The mass sensitivity was determined by counting only the masses on the corresponding test mass subset. The combination of the sensitivity from the test mass subset and the FP rate from the normal data set at the corresponding detection thresholds resulted in a test FROC curve. The training and testing procedure were performed for each cycle of the twofold cross-validation process, thereby generating two test FROC curves. To estimate the overall performance of the CAD system, an average test FROC curve is obtained by averaging the FP rates from the FROC curves of the two mass subsets at the corresponding sensitivities.

Chakraborty *et al.*[32] proposed a JAFROC method and provided software to estimate the statistical significance of the difference between two FROC curves. We employed the JAFROC analysis to evaluate the difference in the FROC curves obtained from the unilateral CAD system and the bilateral CAD system.

III. RESULTS

A. Bilateral feature analysis

Figures 6 and 7 show examples of detection results obtained from the unilateral system and the bilateral system. Figure 6 shows a mass that was initially detected as a mass candidate but was excluded in the false positive reduction steps and was therefore a FN of the unilateral CAD system. The bilateral analysis increased the likelihood score of this mass. It was therefore not excluded in the false positive reduction steps and became a TP in the bilateral CAD system.

Figure 7 shows an example of an FP detected by the unilateral CAD system. The FP was excluded in the bilateral system because it was found to have high symmetry with the tissue in the contralateral breast, as shown in the ROI in Fig. 7(d), by the bilateral analysis.

B. Performance evaluation

In the prescreening process, we obtained a large number of mass candidates on each mammogram. Each mass candidate was paired with a corresponding ROI in the contralateral breast. A total of 3127 and 3402 mass candidates were extracted for training subsets 1 and 2, respectively, which included 98.5% (134/136) and 99.3% (139/140) of the

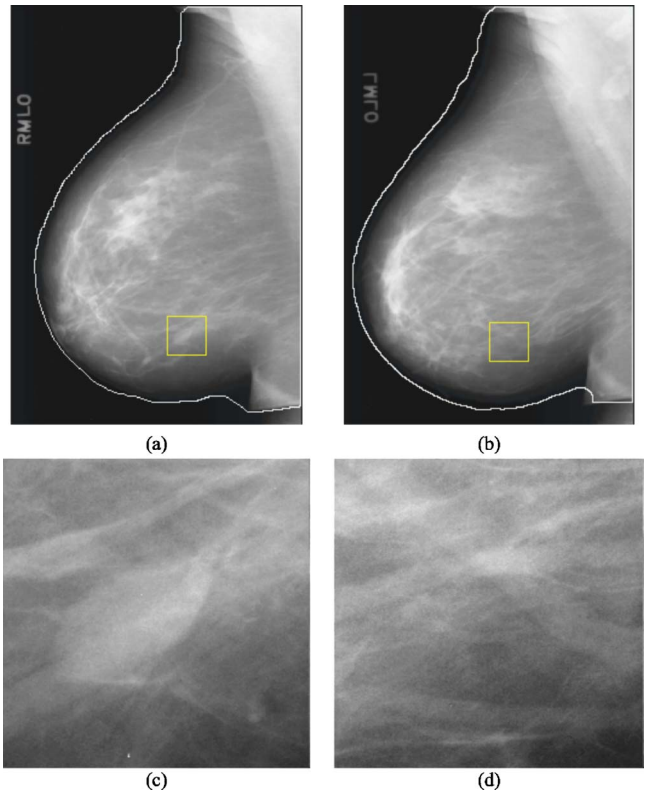


FIG. 6. (a) Mammogram containing a mass marked by the rectangular box. (b) A contralateral mammogram of (a) and the rectangular box is the corresponding ROI of the mass in (a) estimated by the automated regional registration technique. (c) ROI extracted from (a) containing a mass detected at the prescreening stage but excluded at the final stage of the unilateral CAD system. (d) The corresponding ROI in the contralateral breast. Bilateral analysis of this ROI pair increased the likelihood score of the mass which was then detected as a TP in the bilateral CAD system.

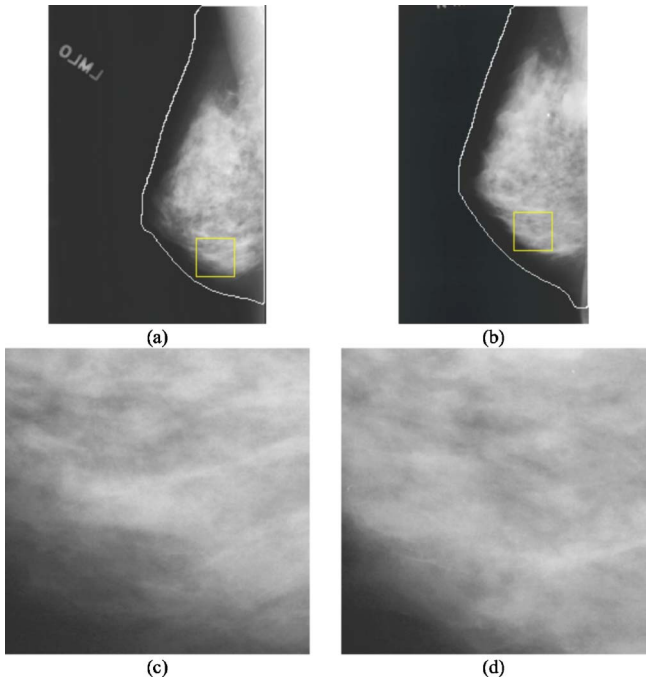


FIG. 7. (a) Mammogram and the rectangular ROI containing a mass candidate. (b) The contralateral mammogram of (a) and the rectangular box is the corresponding ROI of the mass candidate in (a). (c) ROI extracted from (a) containing normal tissue detected at the prescreening stage and included as a FP at the final stage of the unilateral CAD system. (d) The corresponding ROI in the contralateral breast. Bilateral analysis of this ROI pair reduced the likelihood score of the normal tissue which then became a TN in the bilateral CAD system.

masses in the two subsets. The mass candidates in the unilateral mammograms and the ROI pairs from bilateral mammograms in the training subset were used to design the unilateral and bilateral classifiers in each of the twofold cross-validation cycles. The most effective subset of features from the available feature pool was selected for each of the training subsets during the training procedure. For the unilateral LDA classifier, 20 (11 global and 9 local) and 19 (12 global and 7 local) texture features were selected from the two independent training subsets, respectively. For the bilateral LDA classifier, 24 (11 global texture, 9 local texture, and 4 morphological) and 23 (12 global, 8 local, and 3 morphological) features were selected from the two independent training subsets, respectively. The validation A_z values of the LDA classifier during the leave-one-case-out training were 0.846 ± 0.011 and 0.832 ± 0.009, respectively, for the two training subsets using the unilateral LDA classifier, and were 0.862 ± 0.015 and 0.859 ± 0.012, respectively, using the bilateral LDA classifier. The classifiers achieved A_z values of 0.833 ± 0.015 and 0.831 ± 0.011, respectively, for the two test subsets using the unilateral LDA classifier, and 0.853 ± 0.013 and 0.849 ± 0.011, respectively, using the bilateral LDA classifier.

Figure 8 shows the average test FROC curves for the unilateral and bilateral CAD systems after FP reduction with the corresponding trained LDA classifiers when the FP rates were estimated from the test subsets with masses. Figure 9 shows the corresponding results when the FP rates were estimated on the set of no-mass mammograms. Table I summarizes the average FP rates estimated with both the mass and no-mass data sets at several case-based sensitivities.

Because the detection performance of CAD systems on cancer cases is of prime importance, we analyzed the performance of our CAD systems for the subset of cases containing malignant masses. Figure 10 compares the average test FROC curves for the unilateral and bilateral CAD systems on malignant cases only. Figure 11 shows the average test FROC curves for the unilateral and bilateral CAD systems with the sensitivities estimated on malignant cases only and the FP rates estimated on the set of no-mass mammograms. The bilateral CAD system achieved a case-based sensitivity of 70%, 80%, and 85% at average FP rates of 0.35, 0.75, and 0.95 FPs/image, respectively, on the test subset of malignant masses. In comparison to the average FP rates for the unilateral CAD system of 0.58, 1.33, and 1.63 FPs/image, respectively, at the corresponding sensitivities, the FP rates were reduced by 40%, 44%, and 42% with the bilateral symmetry information. Table II summarizes the average FP rates esti-

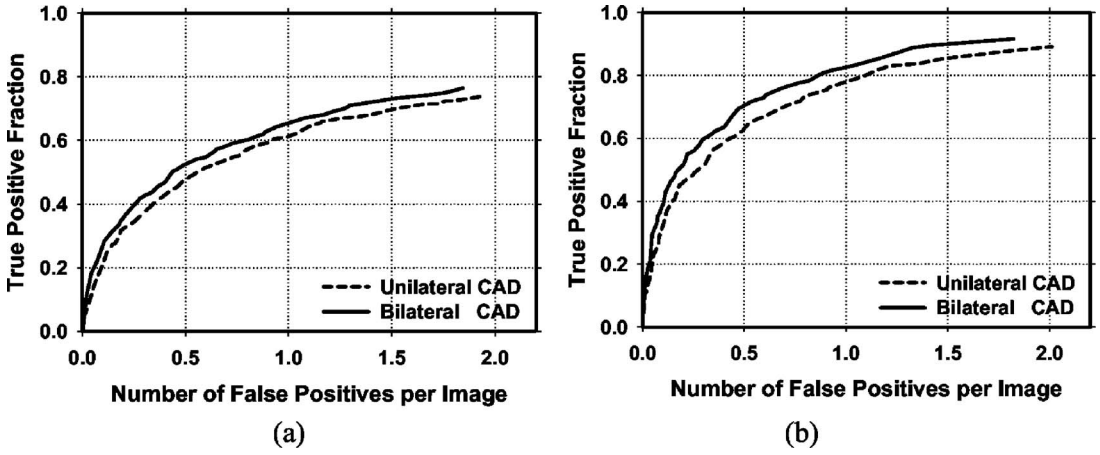


FIG. 8. (a) Image-based and (b) case-based average test FROC curves from the unilateral and the bilateral CAD systems. The FP rates were estimated from detection on mammograms in the test subsets with masses.

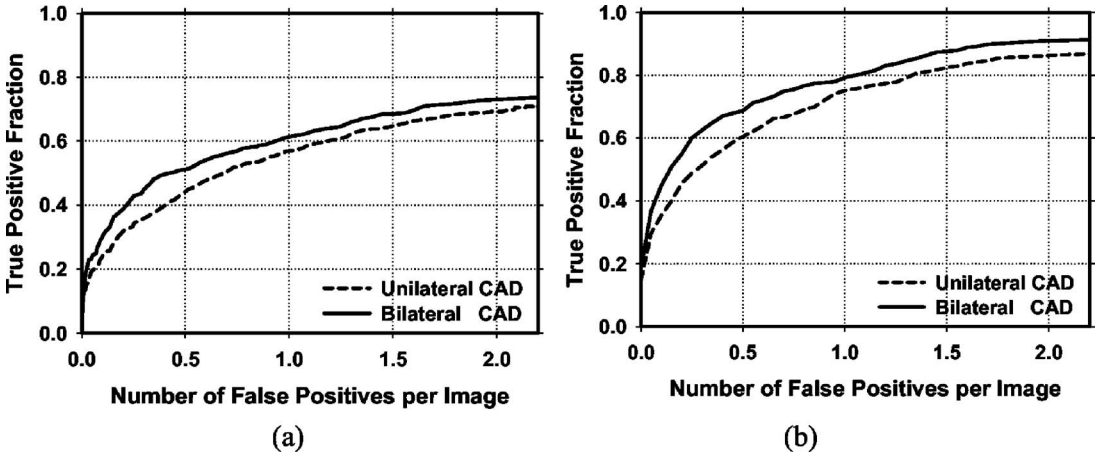


FIG. 9. (a) Image-based and (b) case-based average test FROC curves from the unilateral and the bilateral CAD systems. The FP rates were estimated from detection on mammograms in the no-mass data set.

TABLE I. The average FP reduction rates at case-based sensitivities of 70%, 80%, and 85% for the test subsets when the FP rates were estimated from the mass and no-mass data sets.

	FP rate estimated from mass data set			FP rate estimated from no-mass data set		
	Unilateral CAD	Bilateral CAD	FP Reduction	Unilateral CAD	Bilateral CAD	FP Reduction
70%	0.70	0.53	24%	0.86	0.53	38%
80%	1.10	0.87	21%	1.32	1.04	21%
85%	1.46	1.15	21%	1.72	1.32	23%

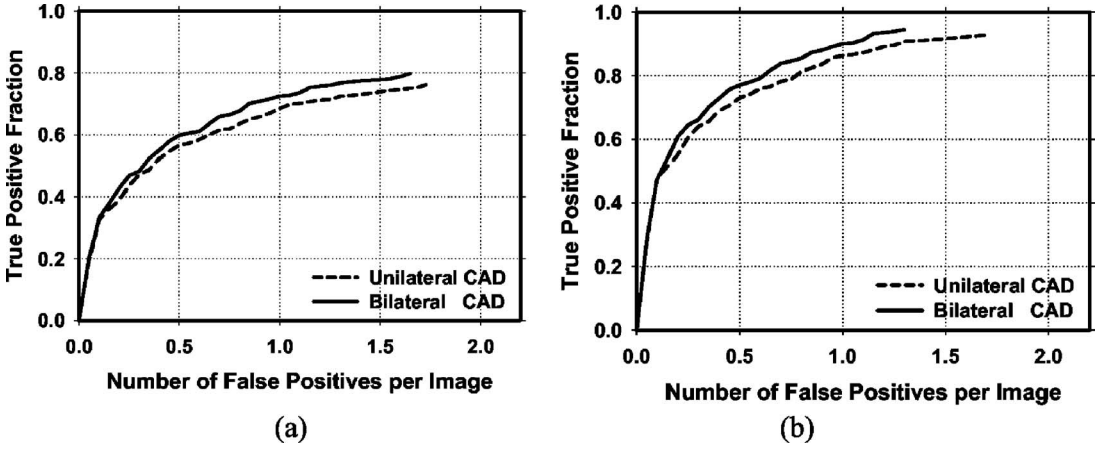


FIG. 10. (a) Image-based and (b) case-based average test FROC curves from the unilateral and bilateral CAD systems for detection on cases with malignant masses only. The FP rates were estimated from in the same data set.

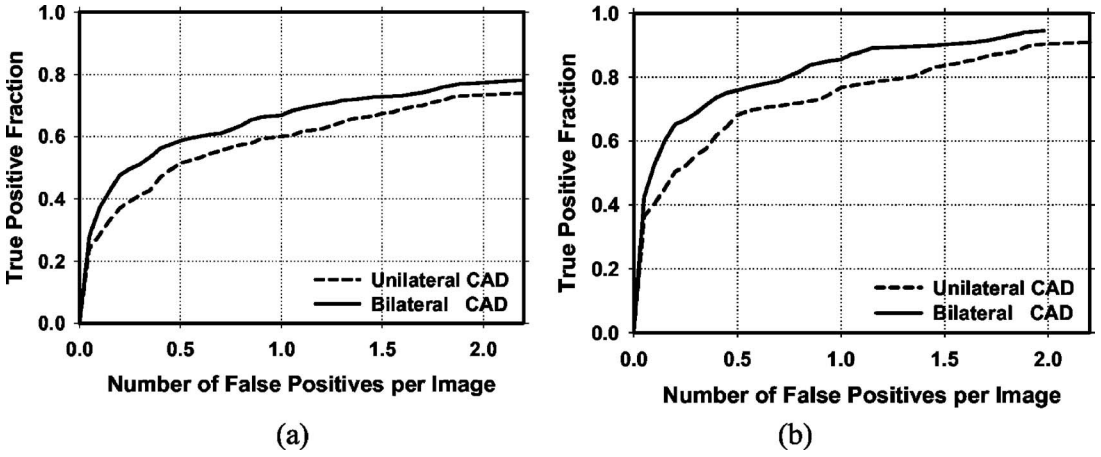


FIG. 11. (a) Image-based and (b) case-based average test FROC curves from the unilateral and bilateral CAD systems for detection on cases with malignant masses only. The FP rates were estimated from the no-mass data set.

TABLE II. The average FP reduction rates for cases with malignant masses at case-based sensitivities of 70%, 80%, and 85% for the test subsets when the FP rates were estimated from the mass and no-mass data sets.

	FP rate estimated from mass data set			FP rate estimated from no-mass data set		
	Unilateral CAD	Bilateral CAD	FP Reduction	Unilateral CAD	Bilateral CAD	FP Reduction
70%	0.43	0.33	23%	0.58	0.35	40%
80%	0.78	0.62	21%	1.33	0.75	44%
85%	0.94	0.78	17%	1.63	0.95	42%

TABLE III. Estimation of the statistical significance in the difference between the FROC performance of the unilateral and bilateral CAD systems on test subsets 1 and 2. The FP rates of the FROC curves were estimated from the no-mass data set: (a) all cases and (b) malignant cases.

(a)	FOM (JAFROC)	
	Test subset 1	Test subset 2
Unilateral CAD	0.52	0.48
Bilateral CAD	0.58	0.51
p value	<0.001	0.008

(b)	FOM (JAFROC)	
	Test subset 1 (malignant only)	Test subset 2 (malignant only)
Unilateral CAD	0.56	0.53
Bilateral CAD	0.61	0.56
p value	0.009	0.003

mated with both the mass and no-mass data sets for cases with malignant masses only at several case-based sensitivities.

The figure-of-merit (FOM) from the output of the JAFROC software is summarized in Table III(a) for all cases and in Table III(b) for malignant cases only. The difference between the FOMs for the unilateral and the bilateral CAD systems was statistically significant ($p<0.05$) for all comparisons.

IV. DISCUSSION

Symmetry between breast structures in bilateral pairs of mammograms is an important feature used by radiologists for mass detection or FP reduction. Similar structures that appear in both right and left mammograms are more likely to be normal tissue than abnormal lesions. Our bilateral analysis translates this radiologists' knowledge to computer vision techniques so that the CAD system can utilize the symmetry of breast tissue on bilateral mammograms to improve detection accuracy. The results of our study show that the bilateral information is an effective technique for reducing FPs.

The bilateral features are important factors affecting the performance of the bilateral LDA classifier. In this study, the bilateral features were derived from features extracted from each pair of ROIs, i.e., the mass candidate and its corresponding ROI, using the maximum-to-minimum ratio strat-

egy as shown in Eq. (1). We also investigated if other strategies, including $BF[i,j]=MC[i,j]/CR[i,j]$, $BF[i,j]=(MC[i,j]-CR[i,j])/MC[i,j]$, and $BF[i,j]=(MC[i,j]-CR[i,j])/[(MC[i,j]+CR[i,j])/2]$, could improve the performance of the bilateral CAD system. It was found that these strategies are not as effective as the maximum-to-minimum ratio. Specifically, among the A_z values of all bilateral features, 72% of those from the latter strategies are lower than those of their corresponding features obtained by Eq. (1). The advantage of using bilateral symmetry measures defined by the maximum-to-minimum ratio can be seen by considering the following example: assuming two ROI pairs that are highly asymmetric, (MC_1, CR_1) and (MC_2, CR_2), in which $MC_1>CR_1$ and $MC_2<CR_2$, their bilateral features derived as the maximum-to-minimum ratio will both be greater than 1. However, the bilateral features obtained from $BF[i,j]=MC[i,j]/CR[i,j]$ will be greater than 1 for (MC_1, CR_1) but smaller than 1 for (MC_2, CR_2). The bilateral measures obtained from $BF[i,j]=(MC[i,j]-CR[i,j])/MC[i,j]$ or $BF[i,j]=(MC[i,j]-CR[i,j])/[(MC[i,j]+CR[i,j])/2]$ will be positive for (MC_1, CR_1) but negative for (MC_2, CR_2). The bilateral feature defined in Eq. (1) therefore describes the asymmetry between the ROI pairs, regardless which ROI has a larger feature value, whereas the other three bilateral features do not consistently provide feature values in the same direction. The maximum-to-minimum ratio approach can thus achieve better performance than the other three strategies.

The corresponding ROI registration is an important procedure in the bilateral analysis. The two breasts of a given patient are not perfectly symmetrical and other factors such as positioning and compression further introduce variability in the symmetry. We investigated the effect of variability in the registered ROI locations on bilateral analysis. For this purpose, the prescreening step of our unilateral CAD system was first applied to the contralateral mammogram to locate the mass candidates. For a given ROI predicted by the registration method on the contralateral mammogram, its location was compared to the ROI locations of these mass candidates by evaluating an overlap ratio, defined as the intersection between the predicted ROI and a mass candidate ROI relative to the area of the smaller ROIs. If the overlap ratio of the predicted ROI with a mass candidate ROI was greater than a chosen threshold, the location of the predicted ROI would be changed to the location of the mass candidate ROI. If the predicted ROI overlapped with more than one mass candidate ROIs, the mass candidate ROI having the largest overlap ratio that exceeded the threshold would be used. We evaluated the effects of this ROI location adjustment for a range of thresholds. It was found that when the overlap ratio threshold was chosen to be about 0.7–0.9, the performance of the bilateral CAD system would have a small but insignificant improvement compared to the bilateral CAD system without the ROI adjustment process. When the overlap ratio threshold was smaller than 0.5, the performance of the bilateral CAD system was degraded. This study indicated that small variability of the predicted ROI location on the contralateral mammogram does not have a strong effect on the performance of the bilateral analysis.

Various registration methods have been attempted for registration of mammograms of the same breast. For example, the warping approach proposed by Sallam *et al.*[33] and the multiple-control-point approach proposed by Vujovic *et al.*[34] Those approaches depended on the identification of corresponding control points. However, there are few, if any, invariant landmarks on mammograms that can be identified automatically because the breast is composed of soft tissue. The projected image of the breast tissue often changes even when the same breast is compressed two different times. It is even more variable between a breast and its contralateral breast. Commonly used rigid or nonrigid registration methods will not be appropriate for this application. We therefore developed the regional registration method for correlation of ROIs on mammograms. Our regional registration method uses the nipple and the distance between the nipple and the ROI center to be the relatively invariant information. The lesion in the target breast is estimated to be located within a band of tissue centered along the arc traced using the nipple-to-lesion distance as the radius and with the origin at the nipple. This method emulates a technique used by many radiologists in identifying corresponding lesions in two-view mammograms or current and prior mammograms. van Engeland *et al.*[35] compared methods for mammogram registration based on breast alignment and linear and nonlinear warping. They concluded that linear warping using mutual information performed better than the other methods. We also performed a study comparing our regional registration method to correlation or mutual information based linear and nonlinear warping methods using a data set of 390 current and prior mammogram pairs.[36] Our results showed that the regional registration method outperformed the warping approaches in identifying corresponding lesions on the mammogram pairs. The localization of symmetric ROIs on the bilateral breasts is similar to the problem of registering ROIs on current and prior mammograms. We therefore adapted the regional registration method to the bilateral analysis in this study.

To implement the bilateral analysis in a practical CAD system, the nipple locations have to be detected automatically. We have previously developed a nipple detection algorithm to determine the nipple location on a mammogram. The algorithm could detect the nipple locations within 1 cm of the manually identified locations in about 70% of the images in the data set used in this study. A large deviation of the nipple location from the true location may affect the regional registration technique in locating the symmetric ROI on the contralateral mammogram, which in turn may degrade the performance of the bilateral analysis of tissue symmetry. We therefore used the manually identified nipple locations in this study in order to develop the bilateral classifier without the influence of other confounding factors. Further work is underway to improve the nipple detection algorithm and to investigate the effect of nipple detection accuracy on the performance of the bilateral system.

The inward nipple projection is often a result of positioning and compression problems so that the nipple is not projected in profile. Since there is not enough information from the two-dimensional projected mammograms to correct for the deformation of the breast, we designed a simple, *ad hoc* correction method to allow the arc drawn using the nipple-to-mass distance as the radius to intersect the breast boundary. In these cases, the breast image on the bilateral mammogram often does not have a similar positioning problem and the difference in the compression of the two breasts may cause large uncertainty in the registration regardless of the correction method. For cases in which both breasts actually have inward nipples and the breast images are similar, our correction method will not cause additional errors because similar correction will be applied to the bilateral mammograms and symmetric ROIs will be identified on the mammograms.

Our motivation of this study is to reduce the FPs of a CAD system for mass detection. A CAD detection system is generally intended for use in screening mammography. At the screening stage, all lesions of concern should be pointed out to radiologists so that the radiologists can judge whether a recall is warranted. If a detection system is trained to mark only the malignant lesions, it may be attempting to play the role of a triage system (alerting radiologists to work up only "malignant" cases) rather than that of a second reader. Furthermore, since computerized lesion detection or characterization on mammograms is not 100% sensitive, it will be confusing to the radiologists whether an unmarked suspicious lesion is missed or it is considered benign by the computer. We believe that computer-aided diagnosis (CADx) may be used in different ways in conjunction with a CAD detection system. For example, the likelihood of malignancy may be estimated by the CADx system and displayed for every detected lesion, and/or a CADx system may be used during diagnostic workup. Either way the CAD system will first alert radiologists to all masses, leaving the assessment of malignancy or benignity to a second stage. We therefore included both malignant and benign masses in the training sets to train the system to detect all masses.

V. CONCLUSIONS

We developed a FP reduction method to improve computerized mass detection on mammograms based on analysis of bilateral information. It was found that the false positives can be reduced by training a new classifier for bilateral features and combining its output score with the unilateral classifier score. The bilateral CAD system achieved a case-based sensitivity of 70%, 80%, and 85% for detection of malignant masses at average FP rates of 0.35, 0.75, and 0.95 FPs/image, respectively, on the test data set. In comparison to the average FP rates for the unilateral CAD system of 0.58, 1.33, and 1.63 FPs/image, respectively, at the corresponding sensitivities, the FP rates were reduced by 40%, 44%, and 42% with the bilateral symmetry information. The improvement in the overall detection accuracy is statistically significant ($p<0.05$) by JAFROC analysis. Our results demonstrate that the bilateral analysis can differentiate the similarity and dissimilarity between tissues at corresponding locations in the bilateral views and is useful for improving the performance of a unilateral CAD system by further reducing the FPs.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA95153 and U. S. Army Medical Research and Material Command Grant Nos. DAMD17-02-1-0214 and W81XWH-1-04-1-0475. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program, and to Dev Chakraborty, Ph.D., for the JAFROC program.

[a]Author to whom correspondence should be addressed. Present address: Department of Radiology, University of Michigan, CGC B2103, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-0904. Telephone: 734-647-8553; Fax: 734-615-5513. Electronic mail: yitawu@umich.edu

[1]American Cancer Society, "Statistics for 2004," www.cancer.org

[2]C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials," Cancer **75**, 1619–1626 (1995).

[3]S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," AJR Am. J. Roentgenol. **171**, 29–33 (1998).

[4]B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," Cancer **91**, 1699–1703 (2001).

[5]L. Tabar *et al.*, "Reduction in mortality from breast cancer after mass screening with mammography," Lancet **325**, 829–832 (1985).

[6]H. C. Zuckerman, in *Breast Cancer, Diagnosis and Treatment*, edited by I. M. Ariel and J. B. Cleary (McGraw-Hill, New York, 1987).

[7]C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists—Findings from a national sample," Arch. Intern. Med. **156**, 209–213 (1996).

[8]R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," Radiology **219**, 192–202 (2001).

[9]J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," J. Natl. Cancer Inst. **95**, 1384–1393 (2003).

[10]L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology **215**, 554–562 (2000).

[11]T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology **220**, 781–786 (2001).

[12]R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," AJR Am. J. Roentgenol. **181**, 687–693 (2003).

[13]S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," Radiology **232**, 578–584 (2004).

[14]H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," Invest. Radiol. **25**, 1102–1110 (1990).

[15]M. A. Helvie *et al.*, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection—A pilot clinical trial," Radiology **231**, 208–214 (2004).

[16]L. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. N. Gurcan, "Analysis of temporal change of mammographic features: Computer-aided classification of malignant and benign breast masses," Med. Phys. **28**, 2309–2317 (2001).

[17]S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Improvement of computerized mass detection on mammograms: Fusion of two-view information," Med. Phys. **29**, 238–247 (2002).
[18]S. van Engeland, S. Timp, and N. Karssemeijer, "Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views," Med. Phys. **33**, 3203–3212 (2006).
[19]F. F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," Med. Phys. **18**, 955–963 (1991).
[20]A. J. Mendez, P. G. Tahoces, M. J. Lado, M. Souto, and J. J. Vidal, "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms," Med. Phys. **25**, 957–964 (1998).
[21]Y.-T. Wu, L. M. Hadjiiski, J. Wei, C. Zhou, B. Sahiner, and H.-P. Chan, "Computer-aided detection of breast masses on mammograms: Bilateral analysis for false positive reduction," Proc. SPIE **6144**, 211–217 (2006).
[22]C. Zhou, H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner, and L. M. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms," Med. Phys. **28**, 1056–1069 (2001).
[23]J. Wei, H.-P. Chan, B. Sahiner, L. M. Hadjiiski, M. A. Helvie, M. A. Roubidoux, C. Zhou, and J. Ge, "Dual system approach to computer-aided detection of breast masses on mammograms," Med. Phys. **33**, 4157–4168 (2006).
[24]N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," Med. Phys. **26**, 1642–1654 (1999).
[25]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," Med. Phys. **32**, 2827–2838 (2005).
[26]R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Trans. Syst. Man Cybern. **SMC-3**, 610–621 (1973).
[27]J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computer-aided image analysis," IEEE Trans. Med. Imaging **12**, 664–669 (1993).
[28]L. Shen, R. M. Rangayyan, and J. E. L. Desautels, "Application of shape analysis to mammographic calcifications," IEEE Trans. Med. Imaging **13**, 263–274 (1994).
[29]S. Mori, H. Nishida, and H. Yamada, *Optical Character Recognition* (Wiley, New York, 1999).
[30]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients," Radiology **224**, 217–224 (2002).
[31]C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat. Med. **17**, 1033–1053 (1998).
[32]D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," Med. Phys. **31**, 2313–2330 (2004).
[33]M. Sallam and K. Bowyer, in *Digital Mammography '96*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (Elsevier, Amsterdam, 1996).
[34]N. Vujovic and D. Brzakovic, "Establishing the correspondence between control points in pairs of mammographic images," IEEE Trans. Image Process. **6**, 1388–1399 (1997).
[35]S. Van Engeland, P. Snoeren, J. Hendriks, and N. Karssemeijer, "A comparison of methods for mammogram registration," IEEE Trans. Med. Imaging **22**, 1436–1444 (2003).
[36]L. M. Hadjiiski, H. P. Chan, B. Sahiner, C. Zhou, M. A. Helvie, and M. A. Roubidoux, "Computerized regional registration of corresponding masses and microcalcification clusters on temporal pairs of mammograms for interval change analysis," in *The 89th Scientific Assembly and Annual Meeting of the Radiological Society of North America*, Chicago, IL, November 30–December 5, 2003, RSNA Program Book 2003 (RSNA, Oak Brook, 2003), p. 389.

Computer-Aided Detection of Breast Masses on Mammograms: Dual System Approach with Two-view Analysis

Jun Wei, PhD
Heang-Ping Chan, PhD
Berkman Sahiner, PhD
Chuan Zhou, PhD
Lubomir M. Hadjiiski, PhD
Marilyn A. Roubidoux, MD
Mark A. Helvie, MD

Department of Radiology

University of Michigan, Ann Arbor

Short title: Two-view dual-system approach for mass detection

Correspondence:
Jun Wei, Ph.D.
Department of Radiology
University of Michigan
1500 E. Medical Center Drive
Med Inn Building C478
Ann Arbor, MI 48109-5842
Telephone: (734) 647-8553
Fax: (734) 615-5513
e-mail: jvwei@umich.edu

ABSTRACT

Purpose: To develop a computer-aided detection (CAD) system that combined a dual system approach with a two-view fusion method to improve the accuracy of mass detection on mammograms.

Methods: We previously developed a dual CAD system that merged the decision from two mass detection systems in parallel, one trained with average masses and another trained with subtle masses, to improve sensitivity without excessively increasing false-positives (FPs). In this study, we further designed a two-view fusion method to combine the information from different mammographic views. Mass candidates detected independently by the dual system on the two-view mammograms were first identified as potential pairs based on a regional registration technique. A similarity measure was designed to differentiate TP-TP pairs from other pairs (TP-FP and FP-FP pairs) using paired morphological features, Hessian feature, and texture features. A two-view fusion score for each object was generated by weighting the similarity measure with the cross correlation measure of the object pair. Finally, a linear discriminant analysis (LDA) classifier was trained to combine the mass likelihood score of the object from the single-view dual system and the two-view fusion score for classification of masses and FPs. A total of 2332 mammograms from 735 subjects including 800 normal mammograms from 200 normal subjects were collected with Institutional Review Board (IRB) approval.

Results: When the single-view CAD system that was trained with average masses only were applied to the test sets, the average case-based sensitivities were 50.6% and 63.6% for average masses on current mammograms and 22.6% and 36.2% for subtle masses on prior mammograms at 0.5 and 1 FPs/image, respectively. With the new two-view dual-system approach, the average case-based sensitivities were improved to 67.4% and 83.7% for average masses and 44.8% and

57.0% for subtle masses at the same FP rates.

Conclusions: The improvement with the proposed method was found to be statistically significant (p<0.0001) by JAFROC analysis.

65

Key words: computer-aided detection, breast mass, false positive reduction

I. **INTRODUCTION**

In screening mammography, two mammographic views, cranio-caudal (CC) and mediolateral oblique (MLO) views, are routinely performed for each breast. During mammographic interpretation, the radiologist combines the information from the two views and evaluates the changes from available prior examinations to confirm true positives (TPs) and to reduce false positives (FPs). It has been reported that screening mammography using two views per breast rather than one view can increase cancer detection sensitivity while decreasing the recall rate [1,2]. Two-view screening mammography has become the most common and standard method for breast cancer screening in developed countries.

Investigators have attempted to implement multiple image techniques in CAD systems to improve the accuracy of lesion analysis on mammograms. Kita *et al.*[3] developed a method to find correspondences between CC and MLO views of the same breast. Their method was based on modeling the deformation of the breast caused by compression in different views. For a data set of 37 lesions, their method could predict the location in the second view with an average minimum distance of 6.78±5.85 mm between the correct position and an epipolar line[3]. Paguerault *et al.*[4] investigated a two-view fusion scheme to improve the performance of a CAD system for mass detection. In their preliminary study, the computer-detected object pairs in two views were first identified by using the distance between the nipple and the detected objects.[5] A trained correspondence classifier was then used to differentiate the TP-TP pairs from other pairs using extracted image features. Finally, a fusion scheme that combined ranking and averaging of the prescreening and correspondence scores was used to estimate a final mass score for each prescreened object. Using 169 pairs of mammograms, they found that the two-view fusion system achieved a significant improvement compared to their single-view CAD system.

In a recent study, Engeland et al. [6] investigated a method in which a two-view classifier was trained with both single-view and two-view features to classify the TP from normal structures instead of training a classifier to differentiate the object pairs. They evaluated the method using 948 cases and found that the method mainly improved the image-based FROC curve in the high specificity range. However, no improvement was found in the case-based FROC curve and they also pointed out that their method may be less relevant when a CAD system is merely used to prompt regions at a high false positive rate. Sahiner *et al*.[7] investigated the use of joint two-view information to improve computerized microcalcification detection. The two-view fusion method was trained and tested on a total of 486 paired mammograms. The improvement in detection with their method was found to be statistically significant for both malignant and benign clusters. Zheng et al [8] proposed a two-view CAD system for masses which aimed to reduce the FP rate on a given sensitivity level. It was found that at a 74.4% case-based sensitivity, their two-view approach reduced the FP rate by 23.7%. Qian et al [9] designed a method for fusing detection results and image features from two views. On a data set of 200 normal mammograms and 200 mammograms containing small (<10 mm) masses, they obtained a significantly improved detection performance when they used their two-view mammogram analysis method. Recently, Velikova et al.[10] proposed a Bayesian network framework that used the dependences between MLO and CC views to obtain a single measure for estimating whether the mammographic view, the breast, and the case contains a cancerous lesion. With the use of the Bayesian network, they obtained a statistically significant improvement compared to single-view analysis for estimating whether the view contains a malignant mass. Furthermore, when the view-based results were combined using logistic regression to estimate whether the breast or the case contains a malignant mass, the improvement was again statistically significant.

The detection of masses on mammograms is a challenging task because the overlapping fibroglandular tissue may mimick a mass or obscures the lesion. Although researchers have devoted extensive efforts to the development of CAD systems for mass detection, the performances of current CAD systems are far from ideal. We have been developing various new techniques to improve the accuracy of mass detection[11,12]. In our previous study, we proposed a dual CAD system approach that combined two mass detection systems in parallel, one was trained with masses of average subtlety and the other with subtle masses. The dual system approach achieved significant improvement in the detection of both average and subtle masses compared to the conventional single system approach[13]. We have also demonstrated the feasibility of a new two-view analysis method for fusion of information from different mammographic views[14]. In this study, our purpose is to further improve the two-view fusion method and to develop a CAD system which combines the dual system approach with the two-view approach. The effectiveness of the new two-view dual CAD system is evaluated with a relatively large data set.

II. MATERIALS AND METHODS

2.1 Image Data Sets

All mammograms in this study were collected retrospectively from patient files of the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. The mammograms were digitized with a LUMISYS 85 laser film scanner with a pixel size of 50μm x 50 μm and 4096 gray levels. The full resolution mammograms were first smoothed with 2x2 box filter and subsampled by a factor of 2, resulting in 100μm x 100μm

images. The images at a pixel size of 100μm x 100μm were used as the input to the CAD system.

Two independent data sets of mammograms were collected for this study: a mass set with biopsy-proven malignant or benign masses and a normal set containing bilateral mammograms. The mass set contained 535 cases with 535 biopsy proven masses in which 345 cases included only current mammograms and 190 cases included both the current and the prior mammograms. 233 of the masses are biopsy proven to be malignant and 302 to be benign. Each case contained two mammographic views (CC view and MLO view or the lateral view). The total number of mammograms in the mass set is 1532 including 1070 current mammograms and 462 prior mammograms in which 35 cases have two prior exams and 3 cases have three prior exams. The true location of each mass was identified independently on each mammographic view by an experienced MQSA-approved radiologist. The masses on the current mammograms are referred to as "average" and the masses on prior exams are referred to as "subtle" because many of those may not show a well-perceived mass even on retrospective review. The normal data set contained 800 mammograms from 200 patients; each case included the CC view and MLO view of both breasts. The normal data set was only used for estimating the FP rate during testing. Figures 1 and 2 show the histograms of mass size and visibility, respectively, for the mass set.

2.2 Methods

Figure 3 shows a schematic of our dual CAD system with two-view analysis. The two-view dual system approach is described in detail below.

A. Dual CAD System Approach

An important purpose of a CAD system is to serve as a second reader to alert radiologists to subtle cancers that may be overlooked. Since the lesions identified on prior mammograms upon retrospective review represent difficult cases that are more likely to be overlooked by radiologists if similar lesions occur on screening mammograms, it is important to improve the sensitivity of the CAD system in detecting these lesions. On the other hand, when a CAD system is applied to a new mammogram in clinical practice, it has to detect breast lesions of all degrees of subtlety effectively. However, it is difficult to train a single CAD system to provide optimal detection for all lesions over the entire spectrum of subtlety because the classifiers have to make compromises to accommodate lesions of a wide range of characteristics.

We have developed a dual system approach and demonstrated that it could improve the overall performance of our CAD system [13]. Briefly, the dual system is composed of two single CAD systems in parallel. The two systems have the same architecture that includes four processing steps: (1) pre-screening of mass candidates, (2) segmentation of suspicious objects, (3) feature extraction and analysis, and (4) FP reduction by classification of normal tissue structures and masses. They were optimized separately by using two different training sets, one contained current mammograms with "average" masses and the other prior mammograms with "subtle" masses. The two data sets did not need to come from the same subjects. After the two single systems were trained separately, they were trained together with a single training set for the dual system information fusion step using an artificial neural network. For an input unknown mammogram, the two systems are applied in parallel and each system estimates a mass likelihood score for every detected object, the trained artificial neural network merges the mass likelihood scores of the two single CAD systems for a given object to differentiate true masses from FPs. The details can be found in the literature[13].

The single-view dual system, described above, constitutes the first stage of the new two-view dual system in the current study. To perform the two-view analysis, a threshold was chosen to retain a small number of the most suspicious objects per mammographic view as input mass candidates to the two-view fusion stage, described next.

B. Two-view Information Fusion

The mass candidates on one view will be paired with mass candidates on the other view based on a regional registration method using geometric criteria. The paired objects will undergo two-view similarity analysis to differentiate TP and FP pairs. The two-view analysis is based on two assumptions: (1) the likelihood of detecting a true mass on both views is higher than that of detecting the same FPs on both views, and (2) the corresponding true masses (TP-TP pair) on two different mammographic views will exhibit higher similarity than that of FP pairs (TP-FP pairs and FP-FP pairs) in terms of morphological features, texture features, and cross correlation.

The key process of our two-view CAD system is the information fusion in which the suspicious objects on different mammographic views are paired together and a unique fusion score is generated for each individual object. Our two-view information fusion scheme consists of four steps: (1) regional registration by using geometric information, (2) estimation of image similarity measure between paired objects using cross correlation, (3) estimation of feature similarity measure by designing a classifier for differentiation of TP-TP pairs from other pairs, and (4) generation of two-view fusion score. Figure 3 shows the block diagram of the two-view information fusion process for suspicious objects on the CC and MLO views of the same breast. Each step is described below in detail.

B.1 Regional registration

Because of the compression of the highly deformable breast and the lack of invariant landmarks in most cases, it is virtually impossible to pinpoint the corresponding locations on different views. We previously developed a regional registration method for locating the approximate locations of corresponding objects on mammograms acquired at different views [4]. From the geometry of the mammographic image acquisition, it is known that an object seen on the CC view can appear only in a limited region in the MLO view, and vice versa. Radiologists at our institution routinely use the nipple-to-object distance (NOD) to estimate the correspondence between objects seen on different views of the same breast. We emulate the radiologists' technique and use the NOD as the geometric matching criterion for initial registration of potential pairs.

The regional registration is performed in a polar coordinate system the origin of which is located at the nipple location. Figure 4 illustrates the process of our regional registration method for a suspicious object on CC view. Using the distance NOD=Rc from the nipple Nc to the center O_{C1} of the object on CC view, an annular region that is bounded by two arcs of radii $Rc \pm \Delta R$ is defined on MLO view with the nipple Nm as the center. The radial width of the annular region $2\Delta R$ was estimated with a large data set to be ±3 cm in our previous study [5]. Any suspicious objects on MLO view that fall within the annular region is paired with the object O_{C1} on the CC view. In this example, O_{m1} and O_{m2} are paired with $O_{C1.}$ After the regional registration process is performed for all suspicious objects detected on the CC view, a number of object pairs that include true mass pairs (TP-TP pairs) and false pairs (FP-TP, TP-FP and FP-FP pairs) are generated.

We developed an automated nipple detection method previously[15] but it did not detect the nipple location correctly in all mammograms. To evaluate the feasibility of the two view

analysis method independent of the nipple detection errors, we used manually identified nipple locations in this study.

B.2 Cross correlation measure

In this step, a template matching approach is used to measure the similarity of the two objects in order to distinguish the truly matched object pairs from the incorrect object pairs. Cross correlation is a popular template matching method. A previous study from our laboratory found that cross correlation was superior to 11 other similarity measures for matching corresponding masses on serial mammograms [16]. In this study, we therefore use cross correlation as the similarity measure to match the same mass appearing on different views. Assume that a mass candidate on the CC view has been paired with several detected objects in the annular region on the MLO view. For a given object pair, the suspicious regions on CC and MLO views are denoted as I^c and I^m, respectively, where the region I^c is a box enclosing the mass candidate detected by the dual CAD system on the CC view and the size of which is determined by the segmentation of the object on this view. The region size is thus varied for each of the candidate object. Because the detected objects may not be centered at the bounding box, a 2 mm $\times$ 2 mm search region is defined with its center at the central location of the paired object on the MLO view. The center of the reference region I^c is placed within the search region and moved one pixel at a time over the entire search region. The cross correlation (r) between I^c and I^m, where I^m is a region with the same size as I^c and centered at each location on the MLO view, is calculated as shown in Eq. (1):

$$r = \frac{\sum_{i=1}^{n}\left[I_i^c - \bar{I}^c\right]\left[I_i^m - \bar{I}^m\right]}{\sqrt{\sum_{i=1}^{n}\left[I_i^c - \bar{I}^c\right]^2}\sqrt{\sum_{i=1}^{n}\left[I_i^m - \bar{I}^m\right]^2}} \quad (1)$$

where I_i^x denotes the i^{th} pixel in the region I^x ($x=c, m$), n is the number of pixels in the reference object region I^c, and

$$\bar{I}^x = \frac{1}{n}\sum_{i=1}^{n} I_i^x \quad (2)$$

The cross correlation measure is defined as the maximum r value among all locations within the search region.

B.3 Two-view similarity classification

We assumed that the features of the same mass on different views will show more similar properties than those of false pairs so that true mass pairs (TP-TP pairs) can be distinguished from false pairs by performing feature classification in the combined space of similarity features.

Three groups of features, morphological features, Hessian features and texture features are extracted from each object. Similarity features are derived as the absolute difference and the mean of the corresponding features of each object pair. These similarity features, in combination with the geometric similarity, i.e., the difference in NOD between the paired objects, formed the feature space for classification of true pairs from false pairs. An LDA classifier was trained to estimate a two-view similarity score for each object pair as detailed in Section D below.

A total of 13 morphological features are extracted as the descriptors of the segmented mass shape. The morphological feature descriptors include the area in terms of the number of pixels in the object, circularity, contrast, convexity, Fourier descriptor, normalized radial length (NRL) mean, NRL area ratio, NRL entropy, NRL standard deviation, NRL zero crossing count, perimeter, perimeter-to-area ratio and rectangularity. The detailed definitions were described in our previous study[17].

Hessian features are derived from the eigenvalues of Hessian matrices in the region of interest (ROI) containing a suspicious object in order to distinguish circular objects from other objects. The Hessian matrix for a 2D image f(x,y) is defined as

$$H_f = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \tag{4}$$

where $f_{xx} = \frac{\partial^2}{\partial x^2} f$, $f_{xy} = f_{yx} = \frac{\partial^2}{\partial x \partial y} f$, and $f_{yy} = \frac{\partial^2}{\partial y^2} f$. To enhance local structures of variable sizes and also reduce the noise, f(*x*,*y*) is convolved with multiscale Gaussian filters having a range of standard deviations (δ_s = 4mm to 10mm) before calculating the Hessian matrices. We designed a response function for mass enhancement at a location (*x*,*y*) and a given scale as

$$R(f(x,y),\delta_s) = \begin{cases} {\lambda_2}^2 / {\lambda_1}^2, & \text{if } \lambda_1, \lambda_2 < 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where λ_1 and λ_2 are the eigenvalues of Eq. (4) with $|\lambda_1| \geq |\lambda_2|$ at the scale with Gaussian filter δ_s. The Hessian feature at a location (x,y) is defined as the maximum value of the response at that location among all scales. Three Hessian features, the Hessian feature at the center location of the ROI (H1), the maximum Hessian feature within the ROI (H2), and the difference between H1 and H2, are calculated for each object.

The texture features are described by the run length statistics (RLS) as follows. The rubber-band straightening transform (RBST) is applied to each object. A band of 60-pixel-wide region around the object margin is transformed to a rectangular image. A gradient magnitude image of the transformed rectangular object margin is derived from Sobel filtering. Five RLS

texture features - short runs emphasis, long runs emphasis, gray level nonuniformity, run length nonuniformity and run percentage - are extracted from the gradient image in both the horizontal and vertical directions, resulting in a total of 10 RLS texture features. Detailed definition of the RBST and the RLS texture features for mammographic masses can be found in the literature [18].

B.4 Generation of two-view fusion score

Since the correspondence of the location of an object projected on different views cannot be determined accurately, several situations will occur. An object on one view may pair with a single object, with multiple objects, or with no object, depending on the number of objects within the annular region on the second view defined for the given object. Each object pair will obtain a similarity score after the LDA classification. We have designed a fusion method to assign a unique score for the suspicious object on the first view from the similarity analysis. The similarity LDA score of the object pair is first weighted by (i.e., multiplied with) the cross correlation measure of the pair. The weighted LDA score is then used as the fusion score for the object if there is only a single object pair. For an object that was paired with multiple objects, the maximum weighted LDA score among all object pairs is chosen as the fusion score for the object. For an object without object pairs, the fusion score is set to be -2.0 as penalty. The value of -2 was chosen because it was slightly smaller than the minimum fusion score obtained in the training set.

C. Two-view system classifier

During this final stage, we have designed a third LDA classifier with two input features, the mass likelihood score from the single-view dual system detection stage and the fusion score from the two-view analysis, to distinguish the mass from normal tissue on each view. The same two-view fusion process is applied to the mass candidates on each view so that each view will

have a set of detected objects with individual scores at the output of this two-view system LDA classifier. The classifier training and testing processes are described below.

D. Training and testing

To train and test the proposed computerized methods, we randomly separated the mass data sets by case into two approximately equal-size independent subsets. Two-fold cross validation was used for training and testing the algorithms. In each cross-validation cycle, we used the training subset for that cycle to select the optimal feature set and train the parameters of the classifiers for the single-view dual system, the two-view similarity analysis, and the two-view dual system. For each classifier, the classification accuracy for the training subset was optimized in terms of the area under the ROC curve, A_z. The single-system LDA classifiers would be trained to combine the multi-dimensional features into the mass likelihood score for each object from the single-view system detection stage, and a neural network classifier was trained to merge the single system scores into a dual system score. The two-view fusion LDA classifier would be trained to combine the multi-dimensional similarity features into a similarity measure for the paired objects. The two-view dual system LDA classifier would be trained to differentiate TPs from FPs.

The LDA classifiers for the single-systems and the two-view similarity analysis were trained with feature selection. Our procedures for feature selection and classifier design have been described in detail elsewhere[11,19,20]. Briefly, feature selection with stepwise LDA[21] and simplex optimization were used to select the best feature subset and reduce the dimensionality of the feature space. The best combination of the stepwise feature selection parameters, including the threshold values for feature entry, feature removal, and tolerance of feature correlation, was first chosen by using a leave-one-case-out resampling method and a simplex optimization

procedure within the training subset. The A_z from the leave-one-case-out testing was used as the figure-of-merit to guide the search for the maximum in the parameter space. Using the best set of parameters and the training subset alone, a final stepwise feature selection was then performed to select a set of features and the weights of the LDA were estimated.

Once the training with one mass subset was completed, the parameters were fixed and applied to the cross-validation test subset. The entire training and testing processes were repeated for the other cross-validation cycle in which the training and test subsets were switched. The set of normal mammograms was not used during training. The trained system from each cycle was applied to the normal set to estimate its FP rate in screening mammograms.

E. Performance analysis

The detection performance of the two-view dual CAD system was assessed by free response ROC (FROC) analysis. An FROC curve was obtained by plotting the mass detection sensitivity as a function of FP marks per image at the corresponding decision threshold. The mass detection sensitivity was determined by the detected masses on the test mass subset whereas the number of FP marks produced by the CAD system was determined by the detected objects on the normal cases only. FROC curves were presented on a per-mammogram and a per-case basis. For image-based FROC analysis, the mass on each mammogram was considered an independent true object. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either mass or both masses on the two views was considered to be a TP detection. Since we used two-fold cross validation method for training and testing, we obtained two test FROC curves, one for each test subset, for each of the conditions (e.g., single-view approach or two-view approach). In order to compare the performance of the single-view and the two-view CAD systems, we applied the jackknife

free-response ROC (JAFROC) method developed by Chakraborty et al.[22] to each pair of the image-based FROC curves obtained with the two systems for the same test subset. To summarize the results for comparison, an average test FROC curve was derived by averaging the FP rates at the same sensitivity along the FROC curves of the two test subsets for each condition.

III. RESULTS

3.1 Single-view dual CAD system

During the first step of our two-view analysis, our previously developed dual CAD system [13] was used as the single-view system to detect mass candidates as input to the later stages. We experimentally chose a criterion of using a maximum of 5 most suspicious mass candidates per image from the single-view detection stage which is a compromise between high sensitivity to retain masses on both views to be paired and the FP rate not being excessively high. With this criterion, the image-based and case-based sensitivities on the current mass set were 88.6% and 95.4%, respectively, while the corresponding sensitivities for the prior mass set were 71.3% and 80.7%, respectively.

3.2 Regional registration

In this study, we used the NOD to register the mass candidates identified by the single view CAD system. Figure 5 showed the histogram of the NOD difference for the same mass which were identified by radiologists on different mammographic views. In our mass set, there were a total of 475 average masses on current mammograms and 107 subtle masses on prior mammograms which could be seen on both views. We used 30 mm as the upper bound to match the object pairs from the same breast and thus the annular region was chosen to have a radial width of ±30 mm. Under this condition, 9 out of 475 average masses and 1 out of 107 subtle

masses were not able to be paired correctly. During the regional registration process, there were a total of 8271 object pairs from the two mass subsets which generated an average of 10.8 object pairs in the CC and MLO views of a breast and 4152 object pairs from the normal data set with an average of 10.4 object pairs in the two views of a breast. After the regional registration, we were able to match only 86.3% (410 of 475) of the mass pairs on current mammograms and 57.9% (62 of 107) of the mass pairs on prior mammograms. Of the average masses, 11.8% (56 of 475) of the misses were caused by either one or both of the masses being missed by the dual CAD system, and only 1.9% (9 of 475) of the average masses could not be matched because the difference in the NODs was larger than 30 mm. For the subtle masses on prior mammograms, the corresponding missed rates were 41.1% (44 of 107) and 0.9% (1 of 107), respectively.

3.3 Two-view similarity classification

For two-view similarity classification, the number of the selected features from the two mass subsets was 6 (Difference of NOD, average of segmented area, average of Hessian output and three average RLS texture features) and 7 (Difference of NOD, average of segmented area, average of Hessian output, difference of NRL entropy and three average RLS texture features), respectively. Figure 6 shows the test ROC curves of the two-view similarity classifier on mass subsets obtained from cross-validation testing with A_z values of 0.87±0.01 and 0.88±0.01, respectively.

3.4 Detection performance comparison

The test FROC curves for average masses on current mammograms are compared in Figure 7. The figures-of-merit (FOM) and the p-values of the difference between pairs of image-based FROC curves under different conditions estimated by JAFROC analysis are tabulated in Table 1. Because of the multiple comparisons, the p-value to achieve statistical significance may

be reduced to 0.002 (=0.05/24) using the conservative Bonferroni correction[23,24]. All paired comparisons achieved statistical significance ($p<0.002$). When the single CAD system was applied to the test sets, the average case-based sensitivities were 50.6% and 63.6% at 0.5 and 1.0 FPs/image, respectively, for the average masses on current mammograms. When the dual CAD system was applied to the test sets, the average case-based sensitivities were improved to 62.1% and 80.1%, respectively, at the same FP rates for the average masses. With the proposed two-view dual system, the average case-based sensitivities were further improved to 67.4% and 83.7%, respectively, at the same FP rates.

The improvement with the proposed approach was also analyzed for the subtle masses on prior mammograms (Figure 8). The FOMs and the p-values of the difference between pairs of image-based FROC curves under different conditions estimated by JAFROC analysis for subtle masses are tabulated in Table 2. The dual system and the two-view dual system have significantly higher ($p<0.002$) detection performances than the single system, whereas the difference between the dual system and the two-view dual system did not achieved statistical significance ($p>0.002$). When the single CAD system was applied to the test subsets, the average case-based sensitivities were 22.6% and 36.2% at 0.5 and 1.0 FPs/image, respectively, for the subtle masses on prior mammograms. When the dual CAD system was applied to the test subsets, the average case-based sensitivities were improved to 41.5% and 55.5%, respectively, at the same FP rates. With the proposed two-view dual system, the average case-based sensitivities for subtle masses were further improved to 44.8% and 57.0%, respectively, at the same FP rates.

Table 1. Estimation of the statistical significance of the difference between the FROC curves for three approaches: the single CAD system, the dual system, and the two-view dual system. The FROC curves with the FP marker rates obtained from the normal data set were compared. The pairs of image-based FROC curves were compared with JAFROC methodology. The figure-of-merit from JAFROC analysis for each curve is shown.

JAFROC Analysis	FOM (Average Masses)			
	All Cases		Malignant Cases	
	Test subset 1	Test subset 2	Test subset 1	Test subset 2
Single system	0.63	0.63	0.58	0.60
Dual system	0.69	0.69	0.68	0.69
P values	<0.0001	<0.0001	<0.0001	<0.0001
Dual system	0.69	0.69	0.68	0.69
Two-view Dual system	0.73	0.72	0.74	0.74
P values	0.0003	0.001	<0.0001	0.0004
Single system	0.63	0.63	0.58	0.60
Two-view Dual system	0.73	0.72	0.74	0.74
P values	<0.0001	<0.0001	<0.0001	<0.0001

Table 2. Estimation of the statistical significance of the difference between the FROC curves for three approaches: the single CAD system, the dual system, and the two-view dual system. The FROC curves with the FP marker rates obtained from the normal data set were compared. The pairs of image-based FROC curves were compared with JAFROC methodology. The figure-of-merit from JAFROC analysis for each curve is shown.

JAFROC Analysis	FOM (Subtle Masses)			
	All Cases		Malignant Cases	
	Test subset 1	Test subset 2	Test subset 1	Test subset 2
Single system	0.42	0.39	0.37	0.32
Dual system	0.48	0.46	0.48	0.45
P values	<0.0001	<0.0001	<0.0001	<0.0001
Dual system	0.48	0.46	0.48	0.45
Two-view Dual system	0.52	0.49	0.52	0.48
P values	0.111	0.078	0.305	0.219
Single system	0.42	0.39	0.37	0.32
Two-view Dual system	0.52	0.49	0.52	0.48
P values	<0.0001	<0.0001	<0.0001	<0.0001

IV. DISCUSSION AND CONCLUSION

We have been developing CAD methods for mass detection on mammograms. We previously designed a dual system approach to improve the overall performance for mass detection[13]. We also conducted a feasibility study of a new two-view analysis method [14]. In this study, we combined these two new approaches into a two-view dual CAD system to further improve its detection accuracy and evaluated its performance in a relatively large data set. Our results indicated that the proposed system could significantly improve the mass detection accuracy in comparison to the single CAD system and the dual CAD system for average masses, whereas the difference in the performances between the two-view dual system and the single-view dual system did not achieve statistical significance for subtle masses.

The improvement achievable with the two-view fusion analysis depends strongly on the sensitivity of the single-view detection stage. If the lesion is missed in the single-view detection, the two-view analysis will not improve the sensitivity. We used the dual-system analysis as the first step in order to detect as many masses as possible (especially for subtle masses) on single views. Although the improvement by dual-system analysis was substantial in comparison with the single CAD system, 110 masses (65 of 475 average masses and 45 of 107 subtle masses) still could not be matched after regional registration. The improvement that was achieved by the two-view analysis was therefore somewhat limited, especially for the subtle masses. For average masses on current mammograms, when we only analyzed the masses which could form TP-TP pairs during regional registration (410 for the average mass set), it was found that the average case-based sensitivities reached 73.4% and 85.7% at an FP rate of 0.5 and 1.0 per image, respectively, with the two-view dual system. Similarly, for the subtle mass set, the average case-based sensitivities reached 67.7% and 80.6% (62 for the subtle mass set) at the same FP rates. It

can therefore be expected that the improvement by two-view analysis will be greater when the single-view detection system can be further improved in the future.

It may be noted that the improvement in detection sensitivity obtained by two-view analysis is different from the apparent increase by case-based FROC analysis. In case-based FROC analysis, a mass is considered to be detected if it is detected either on one view or on two views. With two-view analysis, there is a true improvement in the detection sensitivity, as can be observed from the comparison of the image-based FROC curves. If an additional detected mass is in the other view of a breast for which the mass is already counted as TP in the case-based FROC curve for single-view analysis, this additional detection will not contribute to an improvement in the case-based FROC curve for two-view analysis. This is the reason that the difference between the two case-based FROC curves for the single-view and two-view analysis is smaller than that observed between the two image-based FROC curves. However, we could not conduct a statistical comparison for case-based FROC curves due to the fact that the FPs from the two views might not be independent and a statistical test is not yet available under this situation. Case-based performance is more generally reported by researchers and CAD system manufacturers so that it is more often used for comparing the detection performance between CAD systems. One should note that the actual image-based detection performance of two systems with similar case-based performance can be significantly different. For clinical applications, there is a practical advantage to increase the sensitivity by two-view analysis because radiologists have greater confidence in a lesion being a TP if the same lesion is detected on both views and are less likely to ignore the CAD mark. Dismissing correct CAD marks has been observed to be a major cause of some radiologists not gaining the benefit of using CAD.

In summary, we have developed a two-view dual CAD system to improve computerized detection of breast masses on mammograms. Our results indicate that the proposed CAD system significantly improved the detection performance as estimated by the JAFROC analysis. The improvement by two-view analysis is strongly related to the performance of the single-view detection system. The performance of the two-view dual system can potentially be further improved if the single-view CAD system is improved. We manually identified the nipple locations for the two-view analysis in this study. We will continue to improve the accuracy of our automated nipple detection method[15] so that we can fully automate the two-view analysis in the future.

ACKNOWLEDGEMENTS

This work is supported by U. S. Army Medical Research and Materiel Command grants W81XWH-1-04-1-0475, USPHS grant CA95153, and RX 4300-019-UM (Subcontract of USPHS grant R21/R33 CA 102960 from Geogetown University). The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program and to Dev Chakraborty, Ph.D., for the JAFROC1 program.

REFERENCES

[1]E. Thurfjell, A. Taube, and L. Tabar, "One-view versus 2-view mammography screening - a prospective population-based study," Acta Radiologica 35, 340-344 (1994).

[2]R. Warren, S. Duffy, and S. Bashir, "The value of the second view in screening mammography," British Journal of radiology 69, 105-108 (1996).

[3]Y. Kita, R. P. Highnam, and J. M. Brady, "Correspondence between different view breast X rays using curved epipolar lines," Computer Vision and Image Understanding 83, 38-56 (2001).

[4]S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Improvement of computerized mass detection on mammograms: Fusion of two-view information," Medical Physics 29, 238-247 (2002).

[5]S. Paquerault, B. Sahiner, N. Petrick, L. M. Hadjiiski, M. N. Gurcan, C. Zhou, and M. A. Helvie, "Prediction of object location in different views using geometrical models," Toronto, Canada, June 11-14, 2000.

[6]S. van Engeland and N. Karssemeijer, "Combining two mammographic projections in a computer aided mass detection method," Medical Physics 34, 898-905 (2007).

[7]B. Sahiner, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, C. Paramagul, J. Ge, J. Wei, and C. Zhou, "Joint two-view information for computerized detection of microcalcifications on mammograms," Medical Physics 33, 2574-2585 (2006).

[8]B. Zheng, J. K. Leader, G. S. Abrams, A. H. Lu, L. P. Wallace, G. S. Maitz, and D. Gur, "Multiview-based computer-aided detection scheme for breast masses," Medical Physics 33, 3135-3143 (2006).

[9]W. Qian, D. S. Song, M. S. Lei, R. Sankar, and E. Eikman, "Computer-aided mass detection based on ipsilateral multiview mammograms," Academic Radiology 14, 530-538 (2007).

[10]M. Velikova, M. Samulski, P. J. F. Lucas, and N. Karssemeijer, "Improved mammographic CAD performance using multi-view information: a Bayesian network framework," Physics in Medicine and Biology 54, 1131-1147 (2009).

[11]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," Medical Physics 32, 2827-2838 (2005).

[12]J. Wei, L. M. Hadjiiski, B. Sahiner, H. P. Chan, J. Ge, M. A. Roubidoux, M. A. Helvie, C. Zhou, Y. T. Wu, C. Paramagul, et al., "Computer Aided Detection Systems for Breast Masses: Comparison of Performances on Full-Field Digital Mammograms and Digitized Screen-film Mammograms," Academic Radiology 6, 659-669 (2007).

[13]J. Wei, H.-P. Chan, B. Sahiner, L. M. Hadjiiski, M. A. Helvie, M. A. Roubidoux, C. Zhou, and J. Ge, "Dual system approach to computer-aided detection of breast masses on mammograms," Medical Physics 33, 4157-4168 (2006).

[14]J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, M. A. Helvie, M. A. Roubidoux, C. Zhou, J. Ge, and Y. Zhang, "Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms," Proc. SPIE 6144, 241-247 (2006).

[15]C. Zhou, H.-P. Chan, C. Paramagul, M. A. Roubidoux, B. Sahiner, L. M. Hadjiiski, and N. Petrick, "Computerized nipple identification for multiple image analysis in computer-aided diagnosis," Medical Physics 31, 2871-2882 (2004).

[16]P. Filev, L. M. Hadjiiski, B. Sahiner, H. P. Chan, and M. A. Helvie, "Comparison of similarity measures for the task of template matching of masses on serial mammograms," Medical Physics 32, 515-529 (2005).

[17]B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features," Medical Physics 28, 1455-1465 (2001).

[18]B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," Medical Physics 25, 516-526 (1998).

[19]H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space," Physics in Medicine and Biology 40, 857-876 (1995).

[20]L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. N. Gurcan, "Analysis of Temporal Change of Mammographic Features: Computer-Aided Classification of Malignant and Benign Breast Masses," Medical Physics 28, 2309-2317 (2001).

[21]M. J. Norusis, *SPSS for Windows Release 6 Professional Statistics*, (SPSS Inc., Chicago, IL, 1993).

[22]D. P. Chakraborty, "Validation and statistical power comparison of methods for analyzing free-response observer performance studies," Academic Radiology 15, 1554-1566 (2008).

[23]J. P. Shaffer, "Multiple hypothesis testing," Annu. Rev. Psychol. 46, 561-584 (1995).

[24]T. V. Perneger, "What's wrong with Bonferroni adjustments," British Medical Journal 316, 1236-1238 (1998).

Figure Captions

Figure 1. Distributions of the mass sizes for 535 average masses identified on the current mammograms and 190 subtle masses identified on the prior mammograms. The size for each mass is measured independently as the longest diameter on each mammographic view by an experienced MQSA radiologist. The mean sizes are 15.0±7.7 mm for average masses and 10.9±6.6 mm for subtle masses, respectively.

Figure 2. Histogram of the mass visibility for 1070 average masses by view identified on the current mammograms and 462 subtle masses by view identified on the prior mammograms. The visibility is evaluated on a 10-point rating scale with 1 representing the most visible masses and 10 the most difficult case relative to the cases seen in their clinical practice. Each mass on a mammogram is rated independently by an experienced MQSA radiologist. There are 60 invisible masses on current mammograms and 124 invisible masses on prior mammograms.

Figure 3. Schematic diagram of our dual-system two-view approach for mass detection on mammograms. The system is developed for screening mammography in which all masses, regardless of malignant or benign, are considered positive.

Figure 4. Illustration of the process of our regional registration method for locating potential object pairs on CC and MLO views.

Figure 5. Distributions of the nipple-to-object distance (NOD) differences for the

same mass on different mammographic views identified by radiologists.

Figure 6. The test ROC curves for classification of TP-TP pairs from other pairs on two test mass subsets. The A_z values for the two mass subsets obtained from cross-validation testing were 0.87±0.01 and 0.88±0.01, respectively.

Figure 7. Comparison of the average test FROC curves obtained from averaging the FROC curves of the two independent subsets for average masses on current mammograms. The FP rate was estimated from normal mammograms. (a) Image-based FROC curves, (b) Case-based FROC curves.

Figure 8. Comparison of the average test FROC curves obtained from averaging the FROC curves of the two independent subsets for subtle masses on prior mammograms. The FP rate was estimated from normal mammograms. (a) Image-based FROC curves, (b) Case-based FROC curves.

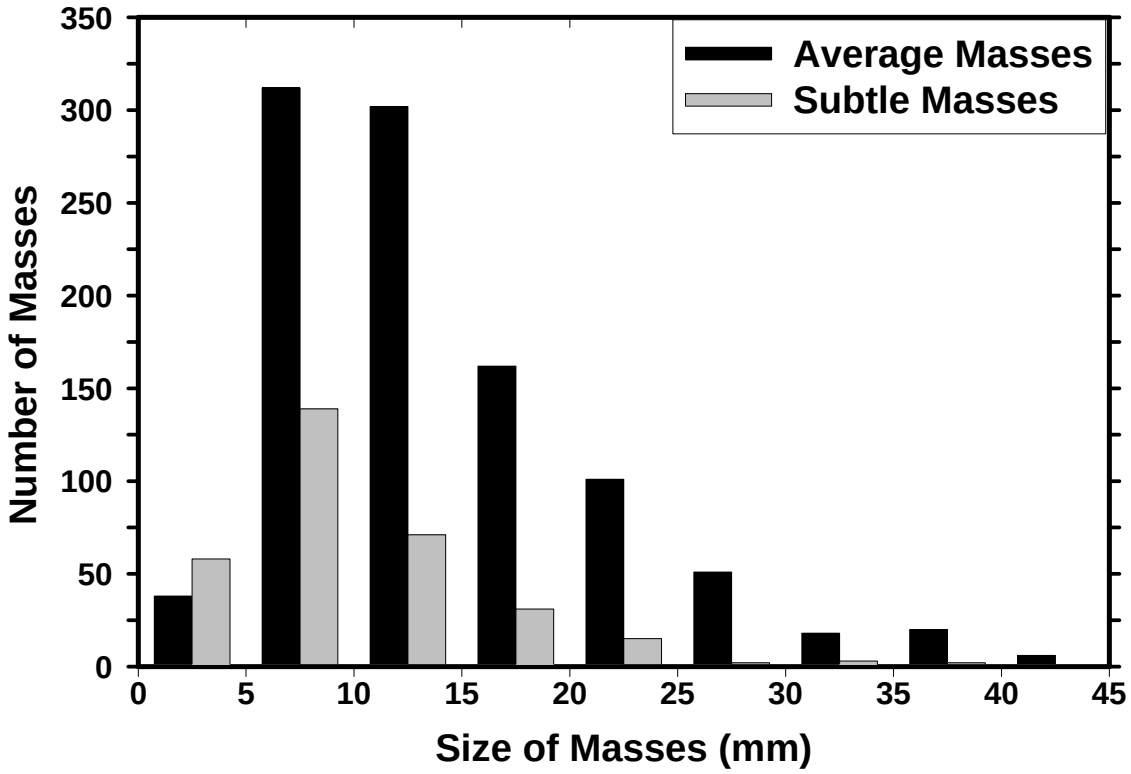


Figure 1. Distributions of the mass sizes for 535 average masses identified on the current mammograms and 190 subtle masses identified on the prior mammograms. The size for each mass is measured independently as the longest diameter on each mammographic view by an experienced MQSA radiologist. The mean sizes are 15.0±7.7 mm for average masses and 10.9±6.6 mm for subtle masses, respectively.

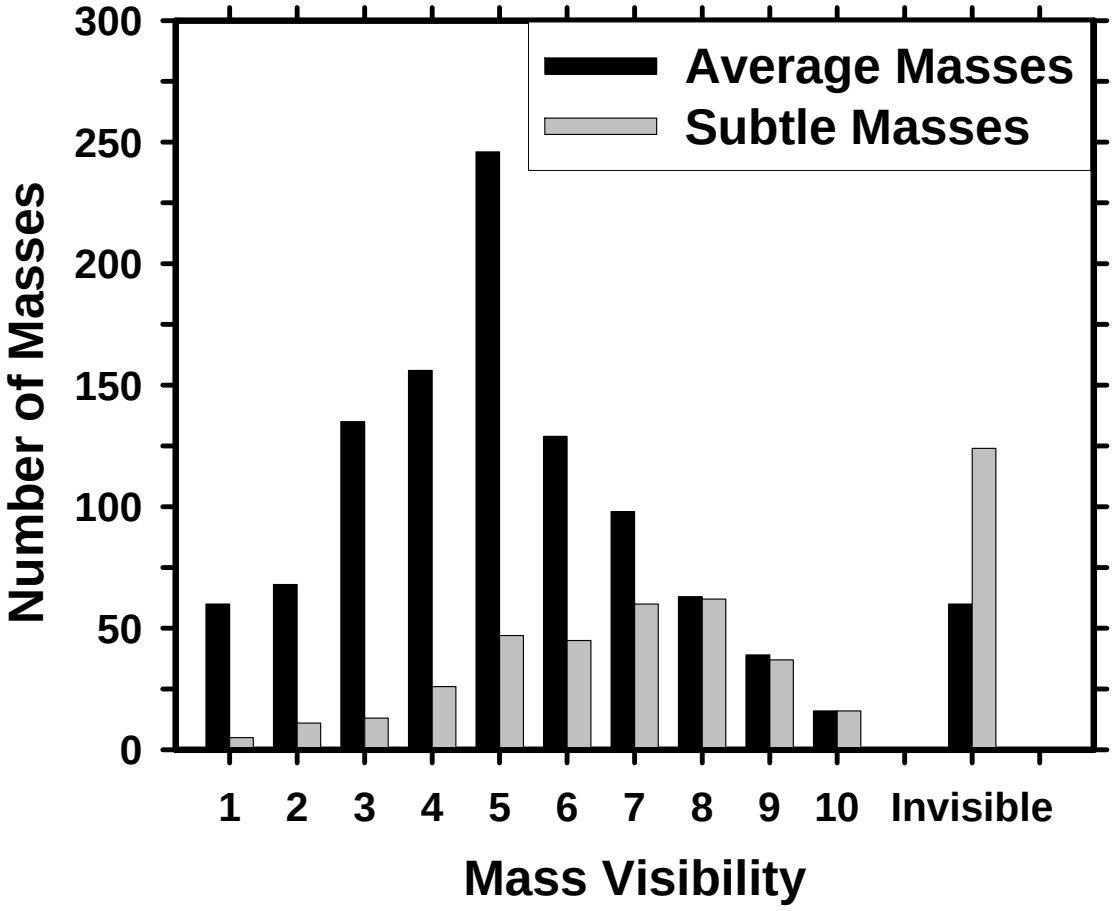


Figure 2. Histogram of the mass visibility for 1070 average masses by view identified on the current mammograms and 462 subtle masses by view identified on the prior mammograms. The visibility is evaluated on a 10-point rating scale with 1 representing the most visible masses and 10 the most difficult case relative to the cases seen in their clinical practice. Each mass on a mammogram is rated independently by an experienced MQSA radiologist. There are 60 invisible masses on current mammograms and 124 invisible masses on prior mammograms.

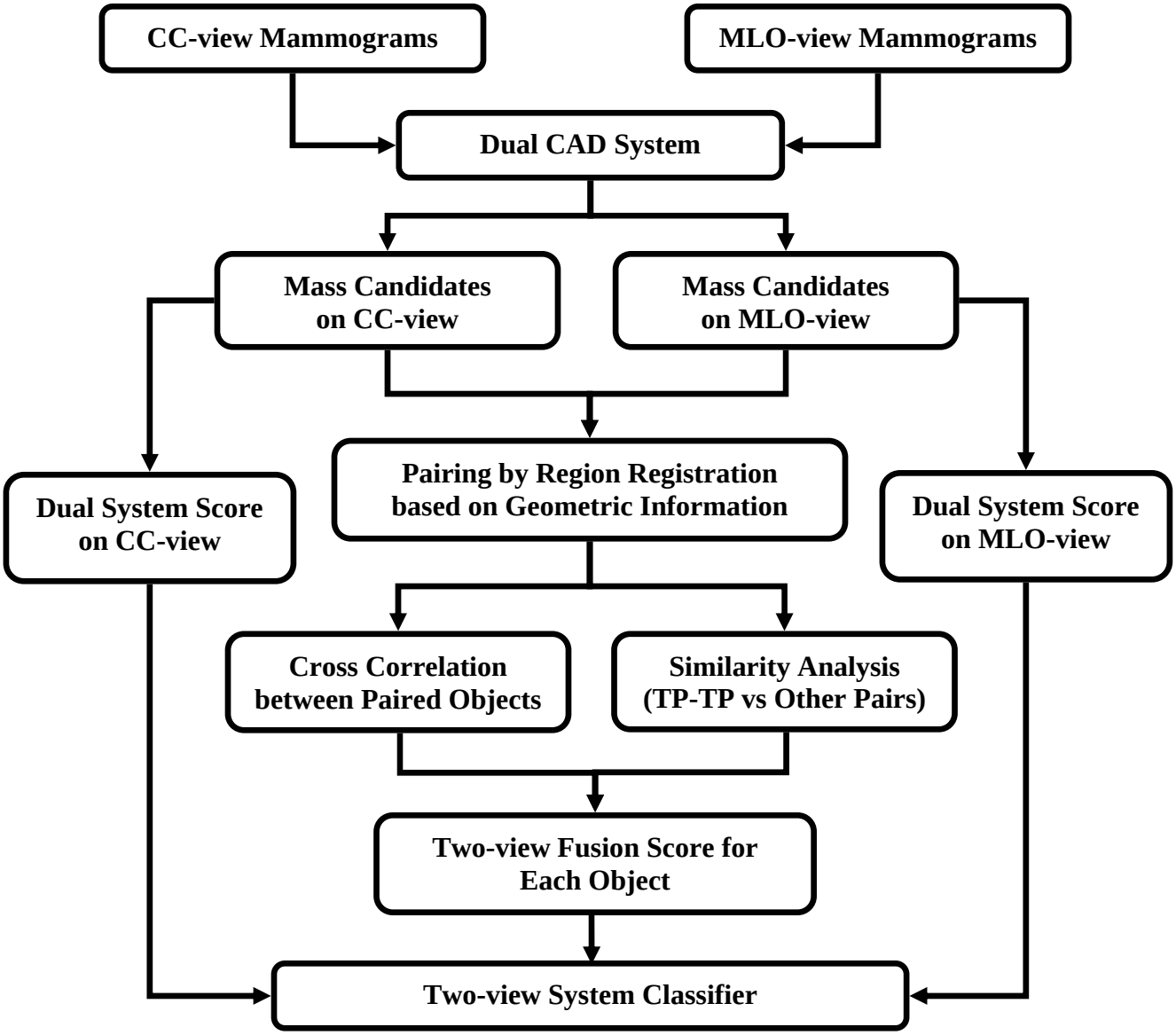


Figure 3. Schematic diagram of our dual-system two-view approach for mass detection on mammograms. The system is developed for screening mammography in which all masses, regardless of malignant or benign, are considered positive.

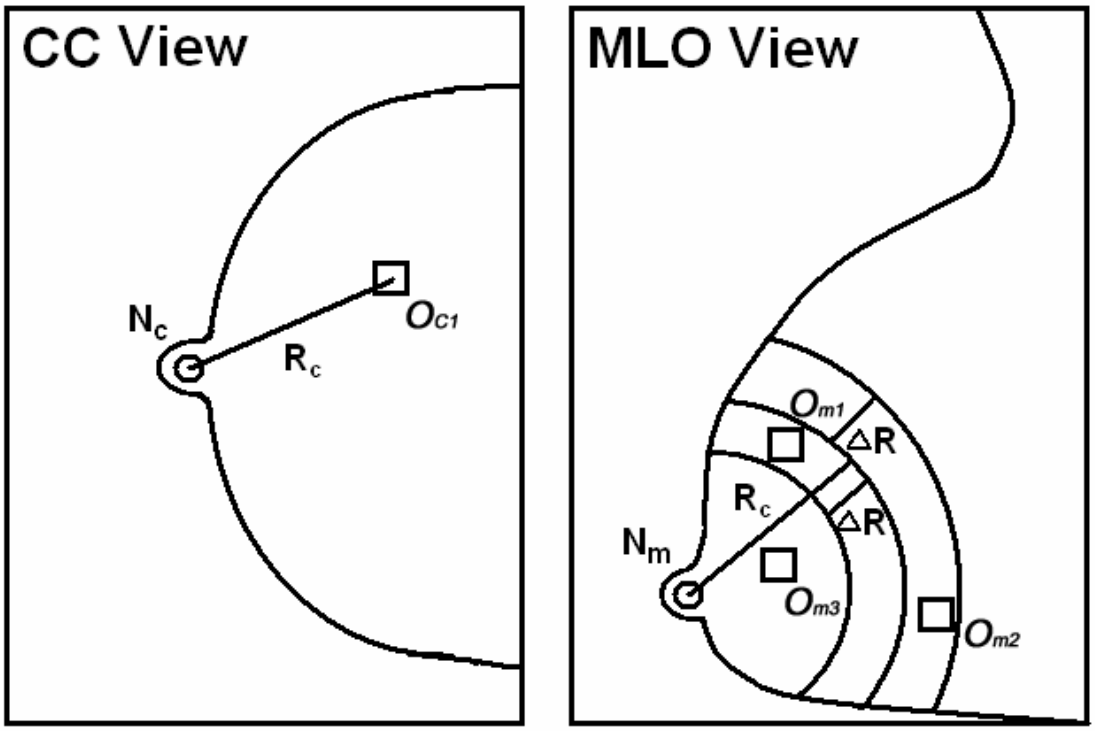


Figure 4. Illustration of the process of our regional registration method for locating potential object pairs on CC and MLO views.

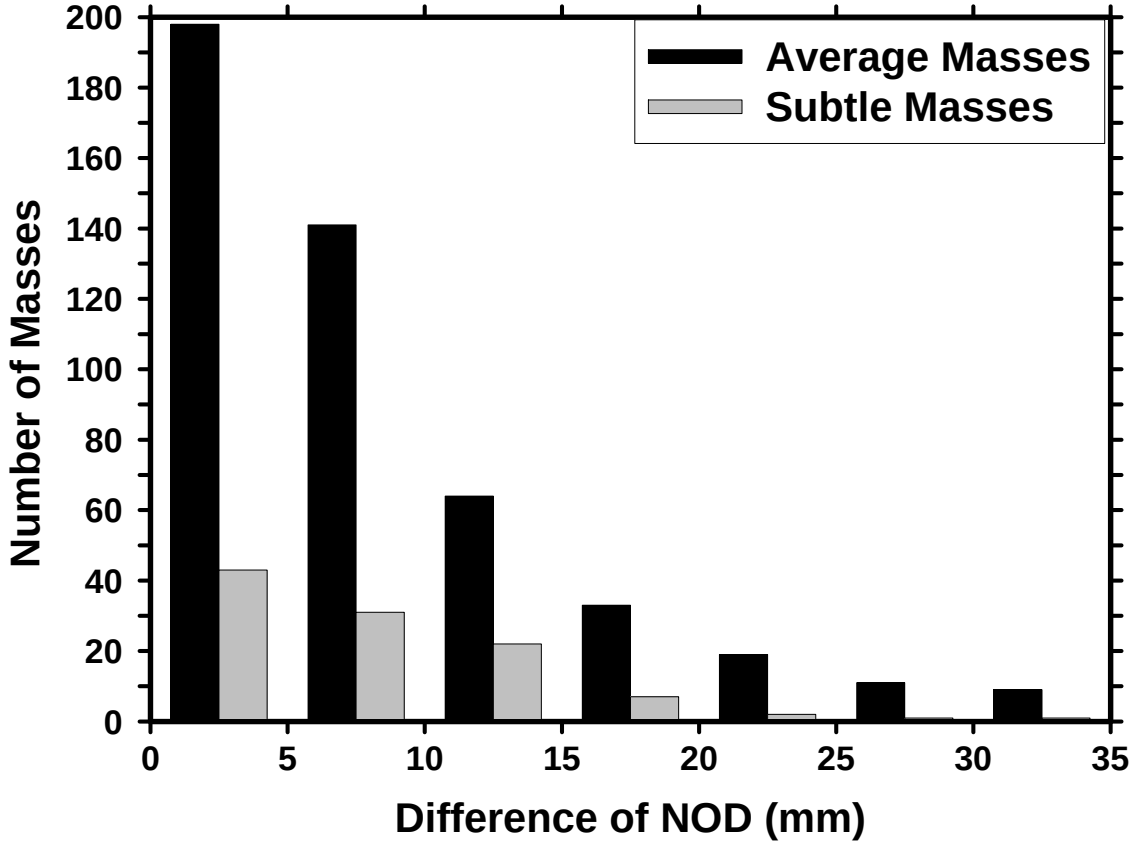


Figure 5. Distributions of the nipple-to-object distance (NOD) differences for the same mass on different mammographic views identified by radiologists.

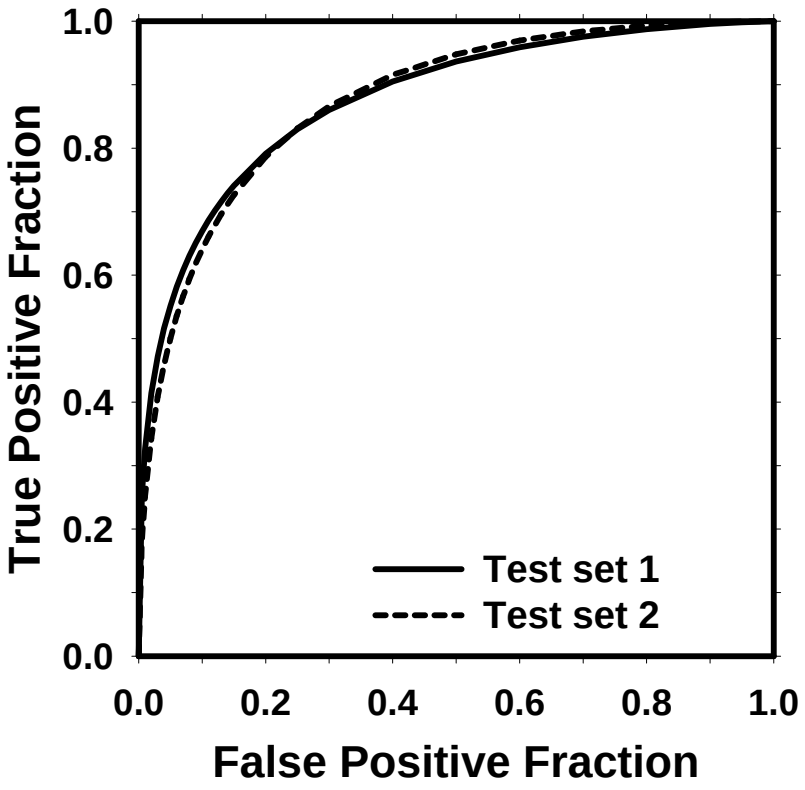
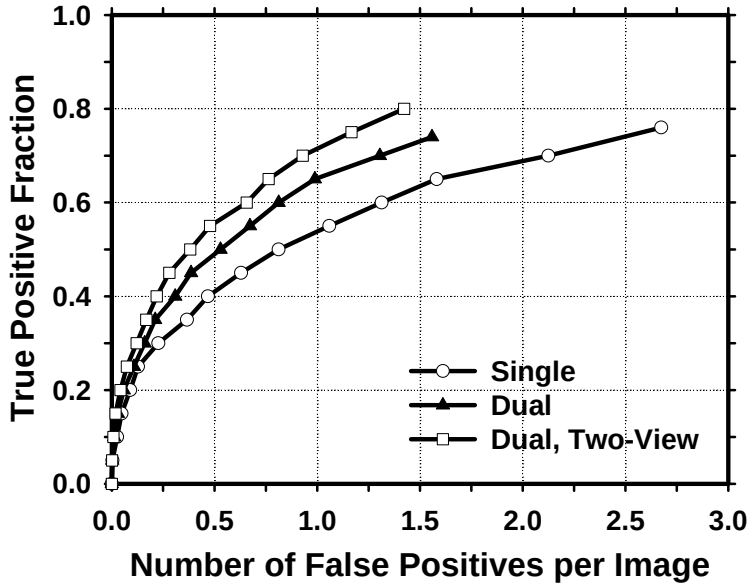
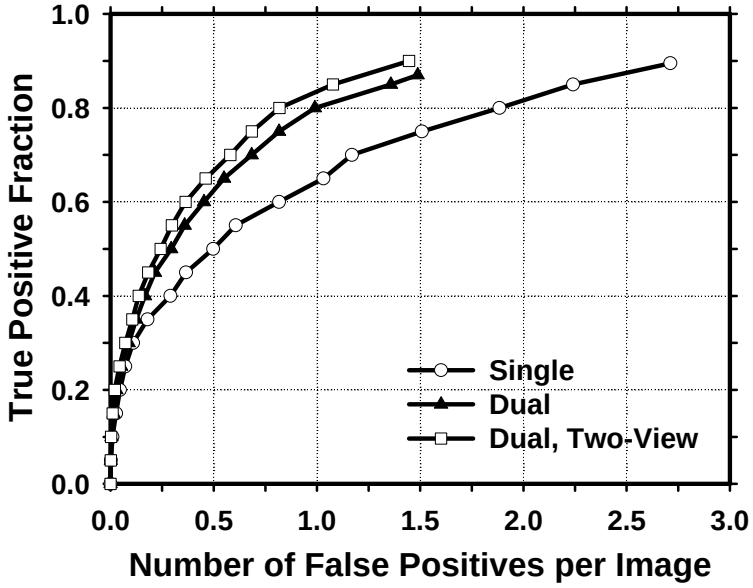


Figure 6. The test ROC curves for classification of TP-TP pairs from other pairs on two test mass subsets. The A_z values for the two mass subsets obtained from cross-validation testing were 0.87±0.01 and 0.88±0.01, respectively.

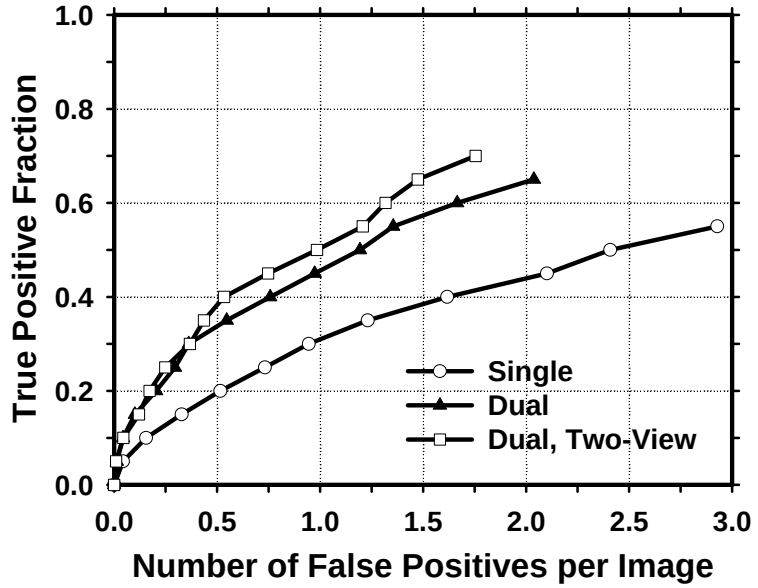


(a)

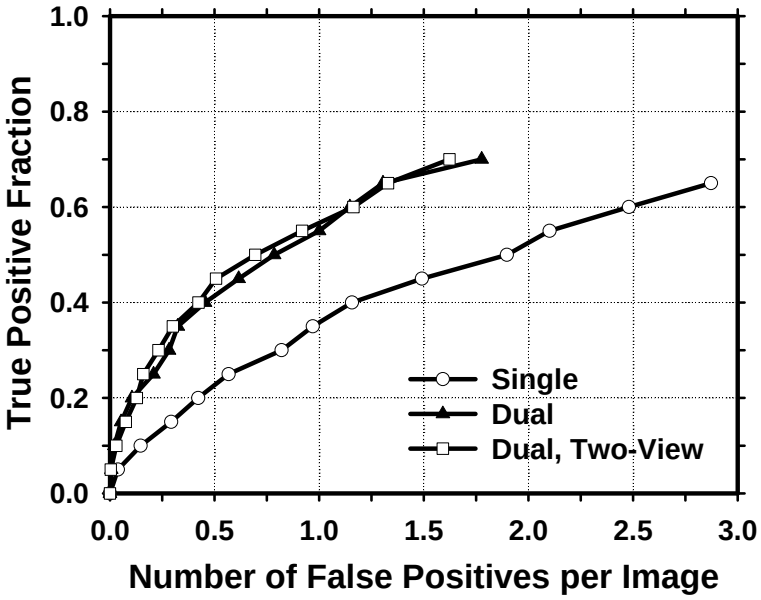


(b)

Figure 7. Comparison of the average test FROC curves obtained from averaging the FROC curves of the two independent subsets for average masses on current mammograms. The FP rate was estimated from normal mammograms. (a) Image-based FROC curves, (b) Case-based FROC curves.



(a)



(b)

Figure 8. Comparison of the average test FROC curves obtained from averaging the FROC curves of the two independent subsets for subtle masses on prior mammograms. The FP rate was estimated from normal mammograms. (a) Image-based FROC curves, (b) Case-based FROC curves.

Computer aided detection of breast masses on full-field digital mammograms: false positive reduction using gradient field analysis

Jun Wei[a], Berkman Sahiner[a], Lubomir M. Hadjiiski[a], Heang-Ping Chan[a],
Nicholas Petrick[b], Mark A. Helvie[a], Chuan Zhou[a], Zhanyu Ge[a]
[a]Department of Radiology, University of Michigan, Ann Arbor
[b]Center of Devices and Radiological Health, U. S. Food and Drug Administration, Rockville, MD

ABSTRACT

Several full-field digital mammography (FFDM) systems have been approved for clinical applications. It is important to develop a CAD system that can easily be adapted to images acquired by FFDM systems from different manufacturers. To develop a CAD system that is independent of the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM image as input and developed a multi-resolution preprocessing scheme for image enhancement. Our CAD system performed prescreening to identify mass candidates, segmented the suspicious structures, extracted morphological and texture features, and then classified masses and normal tissue. In this study, we investigated the use of a two-stage gradient field analysis to identify suspicious masses, and the effectiveness of a new gradient field feature extracted from each suspicious object for false positive (FP) reduction. A data set of 104 cases with 243 images acquired with a GE FFDM system was collected. Most cases had two mammographic views, except for 12 cases that had three views and 1 case with only one view. The data set contained 106 masses. The true locations of the masses were identified by an experienced radiologist. Using free-response receiver operating characteristic (FROC) analysis, it was found that our CAD system achieved a cased-based sensitivity of 70%, 80%, and 88% at 0.8, 1.3, and 1.7 FP marks/image, respectively. The high performance indicated the usefulness of the new gradient field analysis method.

Keywords: Computer-aided diagnosis (CAD), Full field digital mammography (FFDM), Gradient field analysis

1. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 to 55 years of age[1 2-4]. It has been reported that early diagnosis and treatment significantly can improve the chance of survival for patients with breast cancer[3-6]. Although mammography is the best available screening tool for detection of breast cancers, studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially[7-12]. Computer-aided diagnosis (CAD) is considered to be one of the promising approaches that may improve the sensitivity of mammography [13,14]. Computer-aided lesion detection can be used during screening to reduce oversight of suspicious lesions that warrant further work-up. It has been shown that CAD can improve radiologists' detection accuracy significantly [15-17].

Most of mammographic CAD algorithms developed so far are based on digitized mammograms. In the last few years, full-field digital mammography (FFDM) technology has advanced rapidly because of the potential of digital imaging to improve breast cancer detection. Several FFDM systems have become commercially available. We have developed a CAD system for the detection of masses on digitized mammograms in our previous study[18,19]. We are developing a mass detection system for mammograms acquired directly by an FFDM system. In this study, we are investigating the use of gradient field analysis to improve the performance of our mass detection system for FFDMs.

Medical Imaging 2004: Image Processing, edited by
J. Michael Fitzpatrick, Milan Sonka, Proceedings of SPIE Vol. 5370
(SPIE, Bellingham, WA, 2004) · 1605-7422/04/$15 · doi: 10.1117/12.536126

2. MATERIALS AND METHODS

2.1 Materials

The data set we used in this study contained 104 cases with 243 images. All the data were collected with institutional review board (IRB) approval. The raw mammograms in this data set were acquired with a GE FFDM system at a pixel size of $100\mu m \times 100\mu m$ and 14 bits per pixel. Most of the cases had two mammographic views, the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral view, except for 12 cases that had three views and 1 case with only one view. The total number of the masses in this data set is 106, of which 104 were biopsy-proven and 2 were followed up. The true locations of the masses were identified by an experienced breast radiologist.

2.2 Methods

Our CAD system consists of five processing steps: 1) preprocessing by using multi-scale enhancement, 2) pre-screening of mass candidates, 3) identification of suspicious objects, 4) extraction of feature parameters, and 5) classification between the normal and the abnormal regions by using rule-base and linear discrimination analysis (LDA) classifiers. The block diagram for the scheme is shown in Figure 1.

FFDMs generally are pre-processed with proprietary methods before being displayed to readers. The image pre-processing method used depends on the manufacturer of the FFDM system. In an effort to develop a CAD system that is less dependent on specific FFDM systems, the raw digital images are used as input to our system. A preprocessing scheme based on a multi-resolution method[20] has been developed for image enhancement. This scheme consists of three steps. First, the boundary of the breast is detected automatically by using Otsu's method[21]. Second, the Laplacian pyramid is used to decompose the image into multi-scales. A nonlinear weight function is designed to enhance each high-pass component. Finally, the Gaussian pyramid is used to reconstruct the multi-scales. The block diagram for the scheme is shown in Figure 2. An example of an original mammogram and the enhanced mammogram are shown in Figs. 3(a) and 3(b), respectively.

In our previous CAD system developed on digitized screen-film mammograms (SFM), an adaptive density-weighted contrast enhancement (DWCE) filter[18] was developed for prescreening. Although the DWCE filter using the gray level information can identify the suspicious location of masses on mammograms with high sensitivity, the prescreening objects often include a large number of enhanced normal breast structures. In this study, we investigate the use of a new method that combines gradient field information and gray level information to detect the mass candidates on the FFDMs. Gradient field information is commonly used in computer vision or other fields to extract objects or intensity field distributions. Kobatake et al[22] designed a filter, referred to as an iris filter, to calculate the convergence of gradient index around each pixel on SFMs which provided shape information for detection of masses. An extension of the iris filter, referred to as an adaptive ring filter, was developed by Wei et al[23] for detection of lung nodules on chest x-ray images. In this study, we have developed a two-stage gradient field analysis method which does not only use the shape information of masses on mammograms (an extension of the adaptive ring filter) but also incorporates the gray level information by using a region growing technique in the second stage to refine the gradient field analysis.

After prescreening, the suspicious objects are identified by using a clustering based region growing method. Figures 3(c) and 3(d) show the initial detection locations and the grown objects, respectively. For each suspicious object, eleven morphologic features are extracted and rule-based and linear classifiers are trained to remove the detected normal structures that are substantially different from breast masses. Global and local multiresolution texture analysis[24,25] are performed in each region of interest by using the spatial gray level dependence matrix. A new gradient field feature is extracted from each suspicious object and added to the feature space for false positive (FP) reduction. Finally, LDA classification is used to identify potential breast masses. Figure 3(e) shows the final detected objects, and Figure 3(f) shows the locations of these objects superimposed on the mammogram, respectively. Further details of this algorithm can be found in the literature[19].

3. RESULTS

We randomly separated the cases in our data set into two independent equal sized data sets: the training data set contained 52 cases with 120 images and the test data set contained 52 cases with 123 images, respectively. Both the mass detection system with DWCE filtering and that with the two-stage gradient field analysis were trained with the training set, the performance of the two trained systems were compared using the test data set. Our CAD system with the DWCE filter for prescreening of mass candidates achieved a case-based sensitivity of 70% and 80% at 1.4 and 1.7 FP marks/image, respectively. When the DWCE filter was replaced by the gradient field analysis for prescreening, the FP marks/image was reduced to 1.0 and 1.4 at the sensitivity of 70% and 80%, respectively. After the addition of the gradient field feature, the FP was further reduced to 0.8 and 1.3 FP marks/image, respectively, at these sensitivities. Alternatively, the new method can achieve a case-based sensitivity of 88% at 1.7 FP marks/image. Figures 4 and 5 show the comparison of performance by using image-based FROC and case-based FROC curves, respectively.

4. DISCUSSION AND CONCLUSIONS

Several FFDM systems have been approved for clinical applications. It is important to develop a CAD system that can easily be adapted to images acquired by FFDM systems from different manufacturers. In this work, we developed a CAD system that uses the raw FFDMs as the input. Our previous CAD system which was developed on digitized mammograms was adapted to FFDMs by using a new prescreening method that employed gradient field analysis and by retraining the processing parameters. A gradient field feature was extracted for further false positive reduction. The gradient field analysis in combination with the gradient field feature can reduce FPs in mass detection on FFDMs. It was found that our CAD system achieved a cased-based sensitivity of 70%, 80%, and 88% at 0.8, 1.3, and 1.7 FP marks/image, respectively. Further study is underway to improve the CAD system using a larger data set.

ACKNOWLEGMENTS

This work is supported by USPHS grant CA95153, U. S. Army Medical Research and Materiel Command grants DAMD 17-01-1-0326 and DAMD 17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred.

REFERENCES

1. J. R. Harris, M. E. Lippman, U. Veronesi, and W. Willett, "Breast Cancer," *N Engl J Med* **327**, 319-328, 1992.
2. S. H. Landis, T. Murray, S. Bolden, and P. A. Wingo, "Cancer statistics, 1998," *CA Cancer J Clin* **48**, 6-29, 1998.
3. C. Byrne, C. R. Smart, C. Cherk, and W. H. Hartmann, "Survival advantage differences by age: evaluation of the extended follow-up of the Breast Cancer Detection Demonstration Project," *Cancer* **74**, 301-310, 1994.
4. S. A. Feig and R. E. Hendrick, "Risk, Benefit, and Controversies in Mammographic Screening," *In: Syllabus: A categorical Course in Physics Technical Aspects of Breast Imaging*, 119-135, Eds. A. G. Haus and M. J. Yaffe, Radiological Society of North America, Inc, Oak Brook, IL, 1993.
5. C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
6. H. Seidman, S. K. Gelb, E. Silverberg, N. LaVerda, and J. A. Lubera, "Survival experience in the Breast Cancer Detection Demonstration Project," *CA Cancer J Clin.* **37**, 258-290, 1987.
7. B. J. Hillman, L. L. Fajardo, T. B. Hunter, and e. al, "Mammogram interpretation by physician assistants," *AJR* **149**, 907-911, 1987.
8. L. W. Bassett, D. H. Bunnell, R. Jahanshahi, R. H. Gold, R. D. Arndt, and J. Linsman, "Breast cancer detection: one versus two views," *Radiology* **165**, 95-97, 1987.
9. M. G. Wallis, M. T. Walsh, and J. R. Lee, "A review of false negative mammography in a symptomatic population," *Clinical Radiology* **44**, 13-15, 1991.

10. J. A. Harvey, L. L. Fajardo, and C. A. Innis, "Previous mammograms in patients with impalpable breast carcinomas: Retrospective vs blinded interpretation," *American Journal of Roentgenology* **161**, 1167-1172, 1993.
11. R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613-617, 1992.
12. C. Beam, P. Layde, and D. Sullivan, "Variability in the interpretation of screening mammograms by US Radiologists," *Arch Intern Med* **156**, 209-213, 1996.
13. F. Shtern, C. Stelling, B. Goldberg, and R. Hawkins, "Novel technologies in breast imaging: National Cancer Institute perspective," *Society of Breast Imaging Conference*, 153-156, Orlando, Florida, 1995.
14. C. J. Vyborny, "Can computers help radiologists read mammograms?," *Radiology* **191**, 315-317, 1994.
15. H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," *Investigative Radiology* **25**, 1102-1110, 1990.
16. L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* **215**, 554-562, 2000.
17. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
18. N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *IEEE Transactions on Medical Imaging* **15**, 59-67, 1996.
19. N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients.," *Radiology* **224**, 217-224, 2002.
20. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications* **COM-31**, 337-345, 1983.
21. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. System, Man, Cybernetics* **9**, 62-66, 1979.
22. H. Kobatake and S. Hashimoto, "Convergence Index Filter for Vector Fields," *IEEE Transactions on Image Processing* **8**, 1029-1038, 1999.
23. J. Wei, Y. Hagihara, and H. Kobatake, "Detection of Rounded Opacities on Chest Radiographs Using Convergence Index Filter," *ICIAP*, 757-761, Venice, 1999.
24. D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograpms: Multiresolution texture analysis," *Medical Physics* **22**, 1501-1513, 1995.
25. D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis," *Medical Physics* **24**, 903-914, 1997.

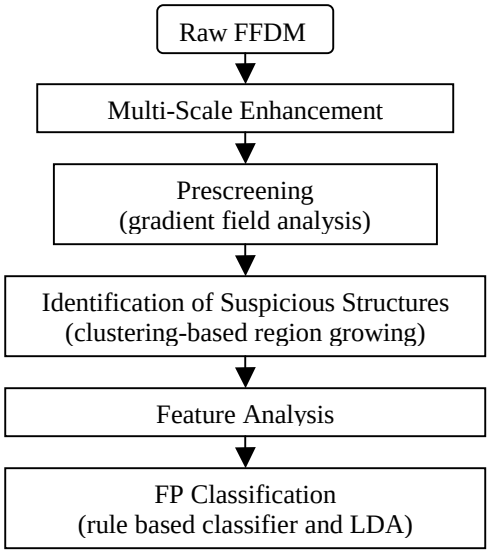


Figure 1: The block diagram of CAD algorithm for mass detection on FFDMs.

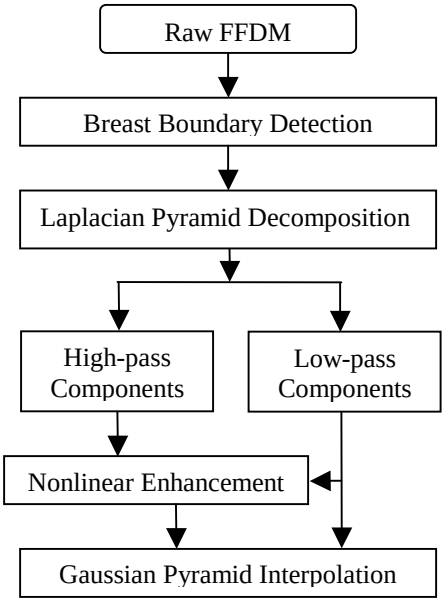
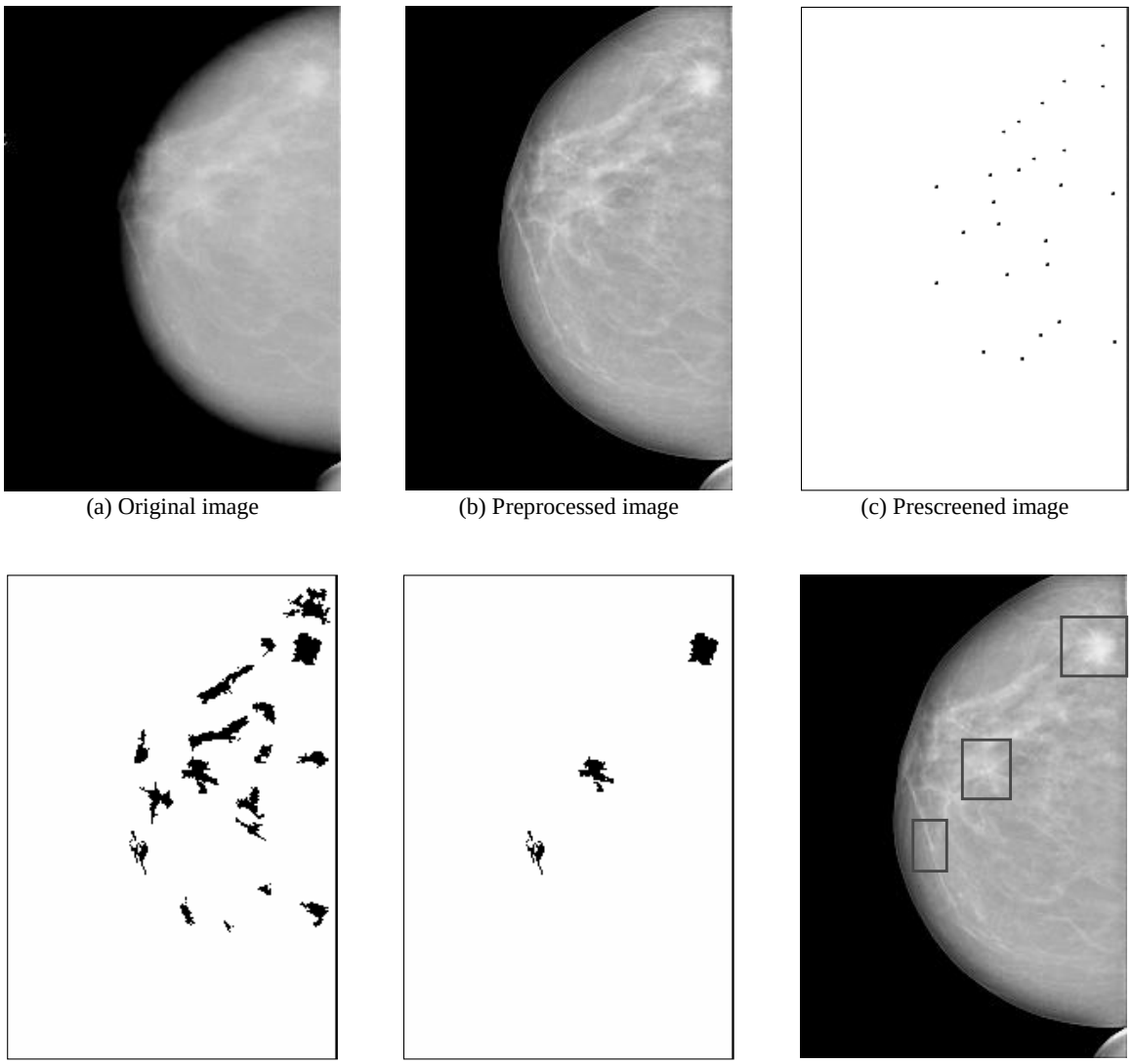


Figure 2: The block diagram for preprocessing of raw FFDM images by multiscale enhancement.



(a) Original image (b) Preprocessed image (c) Prescreened image

(d) Identified suspicious objects (e) Detection result (f) Image with detected objects

Figure 3: An example demonstrating the processing steps with our computer-aided mass detection system.

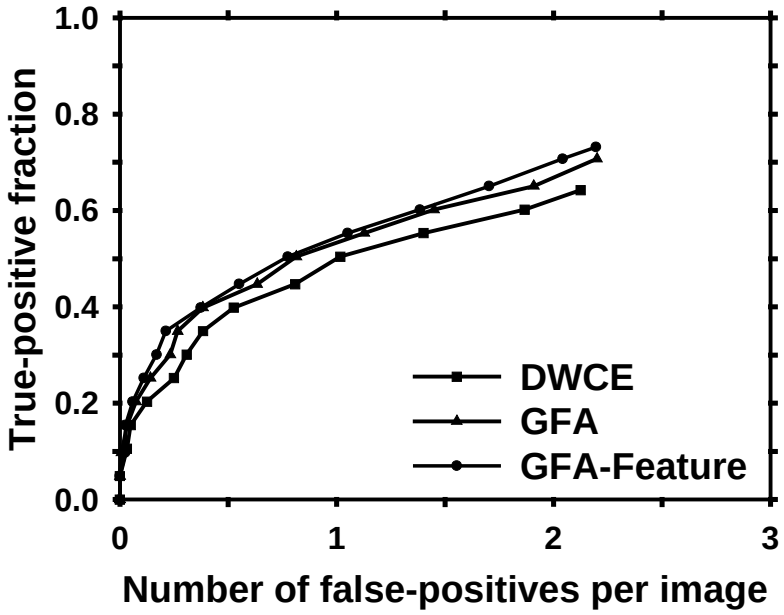


Figure 4: Image-based FROC curves. DWCE: prescreening using DWCE filter. GFA: prescreening using gradient field analysis. GFA-Feature: prescreening using gradient field analysis and the addition of the gradient field feature for FP reduction.

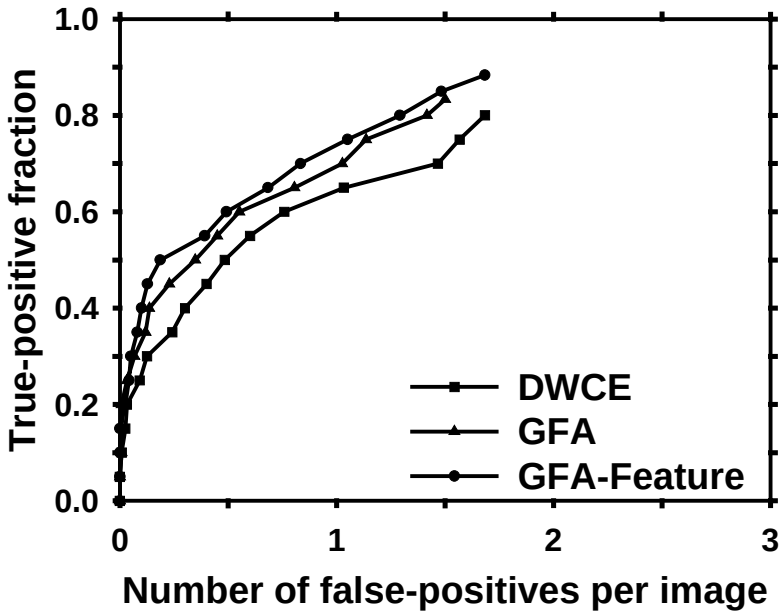


Figure 5: Case-based FROC curves. DWCE: prescreening using DWCE filter. GFA: prescreening using gradient field analysis. GFA-Feature: prescreening using gradient field analysis and the addition of the gradient field feature for FP reduction.

Computer-aided detection of breast masses on mammograms: performance improvement using a dual system

Jun Wei[*a], Berkman Sahiner[a], Lubomir M. Hadjiiski[a], Heang-Ping Chan[a], Mark A. Helvie[a], Marilyn A. Roubidoux[a], Nicholas Petrick[b], Jun Ge[a], Chuan Zhou[a]
[a]Department of Radiology, University of Michigan, Ann Arbor
[b]Center of Devices and Radiological Health, U. S. Food and Drug Administration, Rockville, MD

ABSTRACT

We have developed a computer-aided detection (CAD) system for breast masses on mammograms. In this study, our purpose was to improve the performance of our mass detection system by using a new dual system approach which combines a CAD system optimized with "average" masses with another CAD system optimized with subtle masses. The latter system is trained to provide high sensitivity in detecting subtle masses. For an unknown mammogram, the two systems are used in parallel to detect suspicious objects. A feed-forward backpropagation neural network trained to merge the scores of the two linear discriminant analysis (LDA) classifiers from the two systems makes the final decision in differentiation of true masses from normal tissue. A data set of 86 patients containing 172 mammograms with biopsy-proven masses was partitioned into a training set and an independent test set. This data set is referred to as the average data set. A second data set of 214 prior mammograms was used for training the second CAD system for detection of subtle masses. When the single CAD system trained on the average data set was applied to the test set, the A_z for false positive (FP) classification was 0.81 and the FP rates were 2.1, 1.5 and 1.3 FPs/image at the case-based sensitivities of 95%, 90% and 85%, respectively. With the dual CAD system, the A_z was 0.85 and the FP rates were improved to 1.7, 1.2 and 0.8 FPs/image at the same case-based sensitivities. Our results indicate that the dual CAD system can improve the performance of mass detection on mammograms.

Keywords: computer-aided detection (CAD), mass detection, dual CAD system

1. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 to 55 years of age[1]. It has been reported that early diagnosis and treatment can improve significantly the chance of survival for patients with breast cancer[2-4]. Although mammography is the best available screening tool for detection of breast cancers, studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially[5-7]. Computer-aided detection (CAD) is considered to be one of the promising approaches that may improve the sensitivity of detecting early breast cancer in screening mammography. It has been shown that CAD can increase the cancer detection rate by radiologists both in the laboratory and in clinical practice[8-13].

We have been developing CAD systems for detection and characterization of mammographic masses and microcalcifications. Detection of masses on mammograms is more challenging than detection of microcalcifications because the normal fibroglandular tissue in the breast causes false positives (FPs) by mimicking masses and causes false negatives due to overlapping with the lesions. Therefore, mass detection systems generally have lower sensitivity and higher FP rate than microcalcification detection systems. In this study, we are investigating the effectiveness of a dual system approach for improving the performance of mass detection on mammograms.

[*] jvwei@umich.edu, phone: 734-647-8553, CGC B2103, 1500 E. Medical Center Dr., Ann Arbor, MI 48109-0904

Medical Imaging 2005: Image Processing, edited by J. Michael Fitzpatrick, Joseph M. Reinhardt, Proc. of SPIE Vol. 5747 (SPIE, Bellingham, WA, 2005)
1605-7422/05/$15 · doi: 10.1117/12.595745

2. MATERIALS AND METHODS

2.1 Materials

The data set we used in this study contained 86 cases. Each case included the current mammograms that were obtained before biopsy and the prior mammograms obtained from previous exams. The prior mammograms were used for training the second system because masses on prior mammograms are generally more subtle than those on current mammograms. The subtle mass set does not have to be obtained from the same cases as the average mass set. The current set contained 172 mammograms and the prior set contained 214 mammograms. All data were collected with Institutional Review Board (IRB) approval. The mammograms in this data set were digitized by a Lumiscan laser scanner with a pixel size of $100\mu m \times 100\mu m$ and 12 bits per pixel. All of the current cases had two mammographic views: the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral view. There were 86 biopsy-proven masses in this data set. The true locations of the masses were identified by an experienced MQSA radiologist.

2.2 Methods

In order to improve the performance of our CAD system for detection of subtle masses, we developed a new dual system approach which combines a system trained with "average" masses with another system trained with subtle masses. When the trained dual system is applied to an unknown mammogram, the two CAD systems are used in parallel to detect suspicious objects on a single mammogram. No prior mammogram is needed. The additional FPs from the use of two systems are reduced by feature classification in an information fusion stage. Figure 1 shows the block diagram for the dual system.

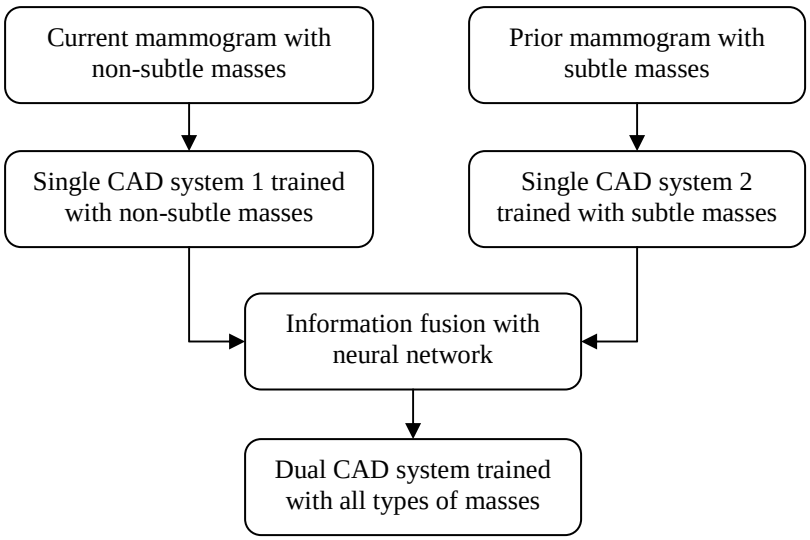


Figure 1. The block diagram of the dual CAD system for mass detection on mammograms.

Our single CAD system consists of five processing steps: 1) digitization, 2) pre-screening of mass candidates, 3) identification of suspicious objects, 4) extraction of feature parameters, and 5) classification between the normal and the abnormal regions by using rule-based and LDA classifiers. The block diagram for the single CAD system is shown in Figure 2. Figure 3 shows an example demonstrating the processing steps with our computer-aided mass

detection system. For the pre-screening stage, we have developed a two-stage gradient field analysis method which uses not only the shape information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis[14,15]. The gradient field analysis was used to determine locations of high convergence of radial gradient in the image. A region of interest (ROI) of 256×256 pixels is then identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method[16] in which the location of high gradient convergence is used as the starting point. Figures 3(b) and 3(c) show the initial detection locations and the grown objects, respectively. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. The objects whose new gradient convergence is lower than 80% of the original value are rejected. After prescreening, the suspicious objects are identified by using a clustering-based region growing method. For each suspicious object, eleven morphological features are extracted. Rule-based and LDA classifiers are trained to remove the detected normal structures that are substantially different from breast masses. Global and local multiresolution texture analysis[17,18] are performed in each ROI by using the spatial gray level dependence matrices at different pixel spacings and angular directions. In order to obtain the best feature subset and reduce the dimensionality of the feature space to design a robust classifier, feature selection with stepwise linear discriminant analysis was applied. Finally, LDA classification is used to identify potential breast masses. Figure 3(d) shows the final detected objects, and Figure 3(e) shows the locations of these objects superimposed on the mammogram.

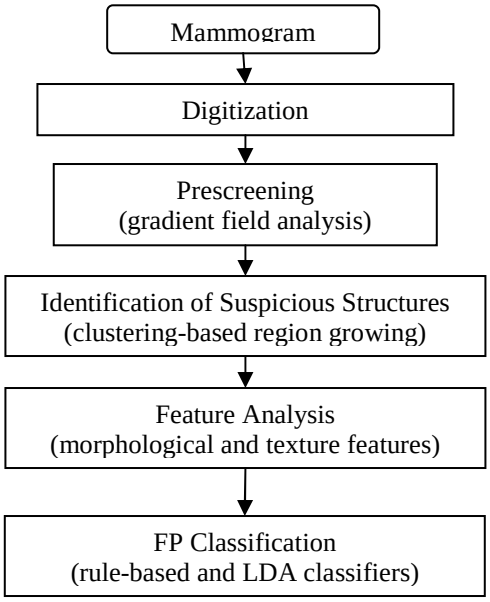
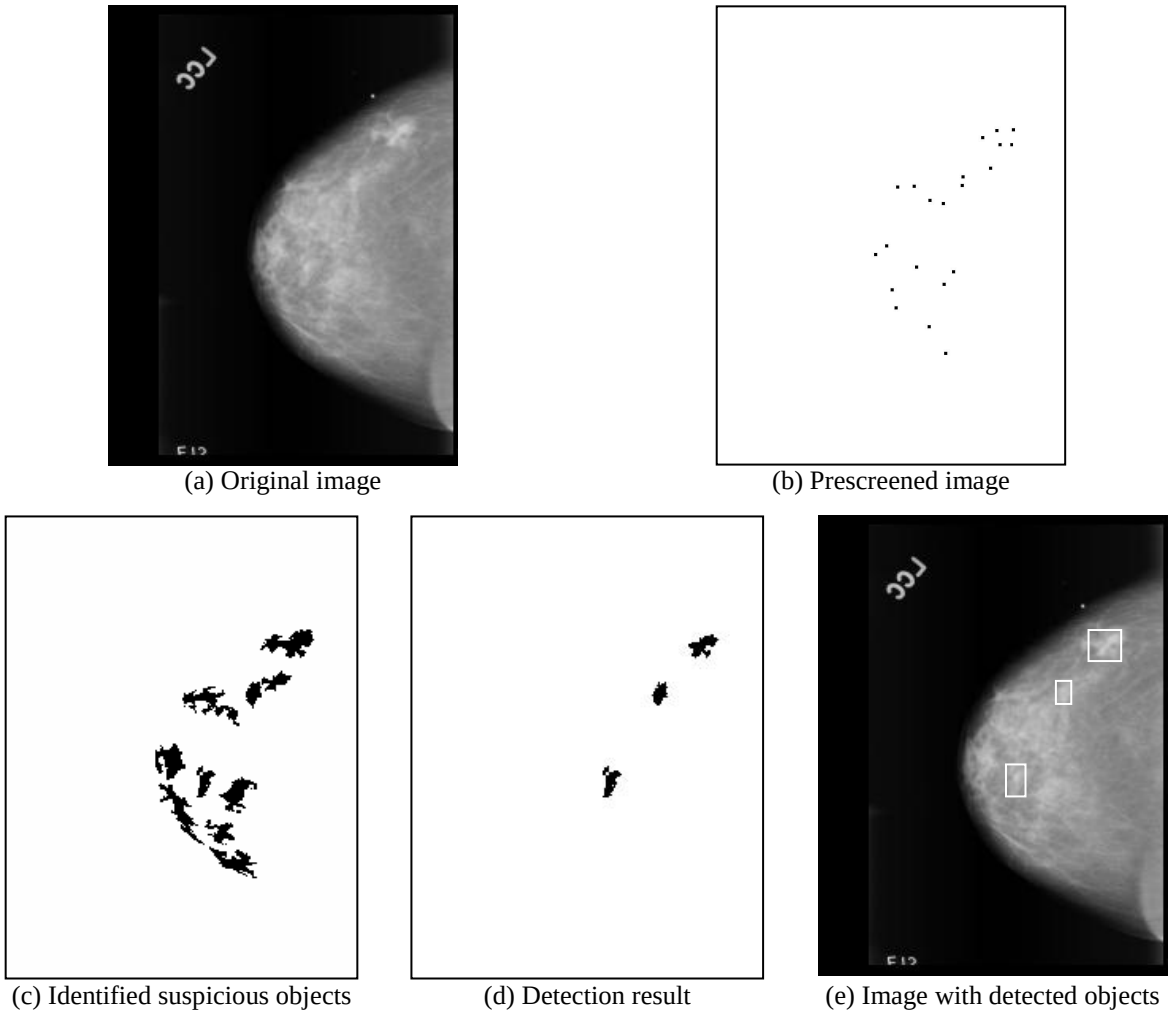


Figure 2. The block diagram of a single CAD system for mass detection on mammograms.

The two single CAD systems were independently trained with the "average" mass set and the subtle mass set, respectively. To merge the information from the two CAD systems, the two LDA discriminant scores from the two CAD systems were used to define a new feature space. A feed-forward backpropagation neural network with 3 hidden nodes was then trained using the LDA feature scores of the training sets as input to differentiate true masses from normal tissue. After the dual CAD system was trained, its performance was evaluated on the independent test set and compared with that of the single CAD system.



(a) Original image (b) Prescreened image

(c) Identified suspicious objects (d) Detection result (e) Image with detected objects

Figure 3. An example demonstrating the processing steps with our single CAD system for mass detection.

3. RESULTS

We randomly separated the cases in our data set into two independent equal sized data sets, each with 43 cases. The training and testing were performed using the cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to 86 masses. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and the detection of either or both masses on the two views was considered to be a true-positive (TP); the sensitivity was thus calculated relative to 43 masses. The average test FROC curve was obtained from averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. When the single CAD system trained on the average data set was applied to the test set, the A_z for FP classification was 0.81 and the FPs/image were 2.1, 1.5 and 1.3 at the case-based sensitivities of 95%, 90% and 85%, respectively. With the dual CAD system, the A_z was 0.85 and the FP rates were improved to 1.7, 1.2 and 0.8 FPs/image at the same case-based

sensitivities. Figure 4 and 5 shows the comparison of the test performance of the single and dual CAD systems by using image-based and case-based average FROC curves, respectively.

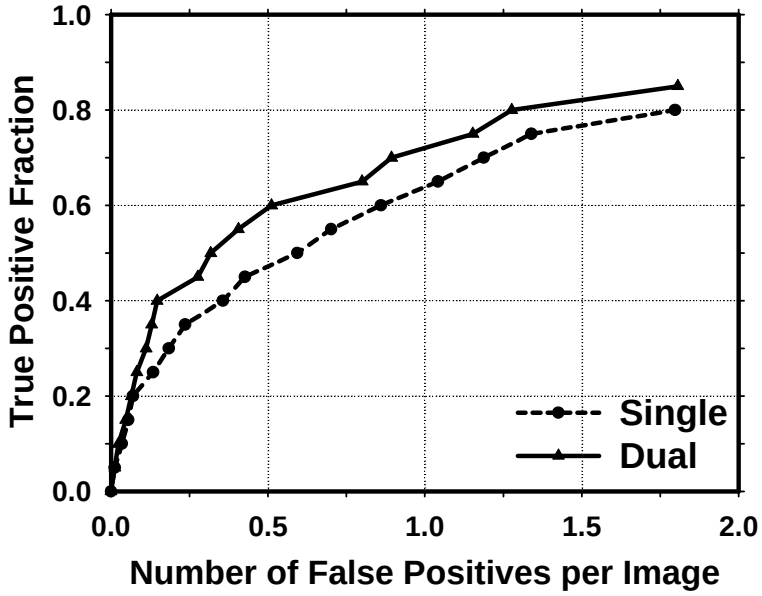


Figure 4. Image-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single: detection by the single CAD system. Dual: detection by the dual CAD system.

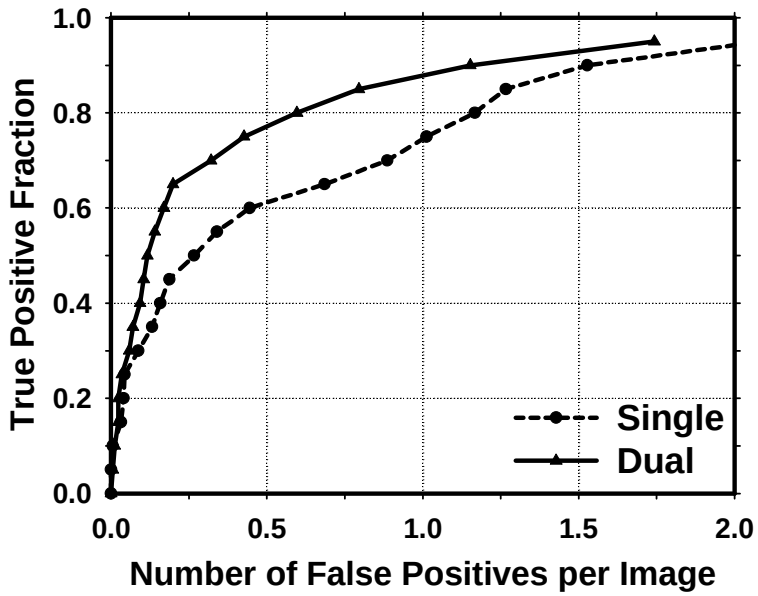


Figure 5. Case-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single: detection by the single CAD system. Dual: detection by the dual CAD system.

4. DISCUSSION AND CONCLUSIONS

We previously developed a CAD system for detection of masses on mammograms. However, we found that it is difficult to train a single system to provide optimal detection for all lesions over the entire spectrum of subtlety. In this study, we developed a dual system which combines a system trained with subtle lesions on prior mammograms and a system trained with masses detected on current mammograms. It was found that the dual CAD system could achieve a higher accuracy than the single CAD system. Further study is underway to optimize the fusion scheme in our dual system.

ACKNOWLEGMENTS

This work is supported by USPHS grant CA95153, U. S. Army Medical Research and Materiel Command grants DAMD17-03-1-0475 and DAMD17-02-1-0214. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

REFERENCES

1. American cancer society, www.Cancer.Org 2005, "Statistics for 2005"
2. C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
3. S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," *Am J Roentgenol* **171**, 29-33, 1998.
4. B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," *Cancer* **91**, 1699-1703, 2001.
5. C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists - Findings from a national sample," *Archives of Internal Medicine* **156**, 209-213, 1996.
6. R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology* **219**, 192-202, 2001.
7. J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," *J. National Cancer Institute* **95**, 1384-1393, 2003.
8. H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," *Investigative Radiology* **25**, 1102-1110, 1990.
9. L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," *Radiology* **215**, 554-562, 2000.
10. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
11. R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," *Am J Roentgenology* **181**, 687-693, 2003.
12. S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," *Radiology* **232**, 578-584, 2004.

13. M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, et al., "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - A pilot clinical trial," *Radiology* **231**, 208-214, 2004.
14. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, C. Zhou, and Z. Ge, "Computer aided detection of breast masses on full-field digital mammograms: false positive reduction using gradient field analysis," *Proc. SPIE Medical Imaging* **5370**, 992-998, 2004.
15. J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Medical.Physics* 2005 (Submitted).
16. N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," *Medical Physics* **26**, 1642-1654, 1999.
17. D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograpms: Multiresolution texture analysis," *Medical Physics* **22**, 1501-1513, 1995.
18. D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis," *Medical Physics* **24**, 903-914, 1997.

Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms

Jun Wei, Berkman Sahiner, Lubomir M. Hadjiiski, Heang-Ping Chan, Mark A. Helvie,
Marilyn A. Roubidoux, Chuan Zhou, Jun Ge, Yiheng Zhang
Department of Radiology, The University of Michigan, Ann Arbor, MI 48109

ABSTRACT

We are developing a two-view information fusion method to improve the performance of our CAD system for mass detection. Mass candidates on each mammogram were first detected with our single-view CAD system. Potential object pairs on the two-view mammograms were then identified by using the distance between the object and the nipple. Morphological features, Hessian feature, correlation coefficients between the two paired objects and texture features were used as input to train a similarity classifier that estimated a similarity scores for each pair. Finally, a linear discriminant analysis (LDA) classifier was used to fuse the score from the single-view CAD system and the similarity score. A data set of 475 patients containing 972 mammograms with 475 biopsy-proven masses was used to train and test the CAD system. All cases contained the CC view and the MLO or LM view. We randomly divided the data set into two independent sets of 243 cases and 232 cases. The training and testing were performed using the 2-fold cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. The average test FROC curve was obtained from averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. At the case-based sensitivities of 90%, 85% and 80% on the test set, the single-view CAD system achieved an FP rate of 2.0, 1.5, and 1.2 FPs/image, respectively. With the two-view fusion system, the FP rates were reduced to 1.7, 1.3, and 1.0 FPs/image, respectively, at the corresponding sensitivities. The improvement was found to be statistically significant ($p<0.05$) by the AFROC method. Our results indicate that the two-view fusion scheme can improve the performance of mass detection on mammograms.

Keywords: computer-aided detection, two-view fusion, mass detection, AFROC analysis

1. INTRODUCTION

Breast cancer is one of the leading causes of cancer mortality among women[1]. There is considerable evidence that early diagnosis and treatment significantly improves the chance of survival for patients with breast cancer [2-5]. Although mammography has a high sensitivity for detection of breast cancers when compared to other imaging modalities, studies indicate that radiologists do not detect all carcinomas that are visible upon retrospective analyses of the images[6-11]. It has been shown that computer-aided detection (CAD) can improve the sensitivity of mammography in prospective clinical trials[12-15]. CAD is thus a viable cost-effective alternative to double reading by radiologists.

The mass detection systems to-date generally employed a single-view detection approach using various techniques for prescreening of mass candidates and classification of true and false positives[16-25]. We have been developing CAD systems for detection of mammographic masses on full field digital mammograms (FFDMs)[25] and screening film mammograms (SFMs)[22]. Our previous study[23] showed that two-view fusion method can improve the performance of a CAD system for mass detection on mammograms. In this study, our purpose is to improve the performance of the two-view information fusion method and to test our method in a relatively larger data set.

2. MATERIALS AND METHODS

2.1 Materials

Medical Imaging 2006: Image Processing, edited by Joseph M. Reinhardt, Josien P. W. Pluim,
Proc. of SPIE Vol. 6144, 614424, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.654593

All mammograms in this study were collected from patient files in the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. The mammograms were digitized with a LUMISYS 85 laser film scanner with a pixel size of 50μm×50μm and 4096 gray levels. The scanner was calibrated to have a linear relationship between gray levels and optical densities (O.D.) from 0.1 to greater than 3 O.D. units. The nominal O.D. range of the scanner is 0–4. The full resolution mammograms were first smoothed with a 2×2 box filter and subsampled by a factor of 2, resulting in images with a pixel size of 100μm×100μm. These images were used for the input of our CAD system. The data set we used in this study contained 475 cases, of which 464 cases had the two-view mammograms (the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral view) and 11 cases had four-view mammograms, resulting in a total of 972 mammograms. All mammograms were obtained before biopsy. There were 475 biopsy-proven masses in this data set.

2.2 Methods

2.2.1 Single-view System Overview

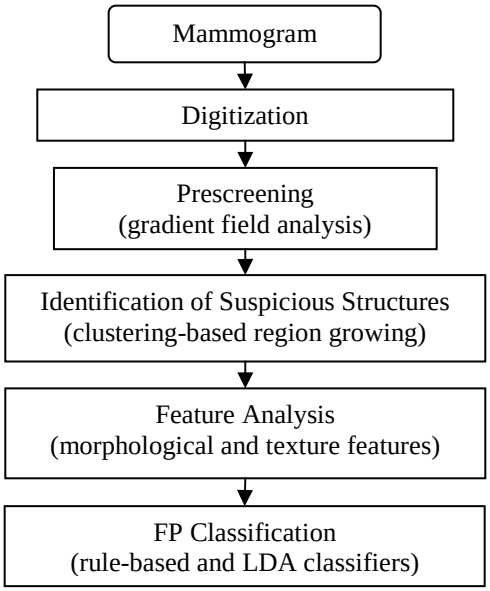
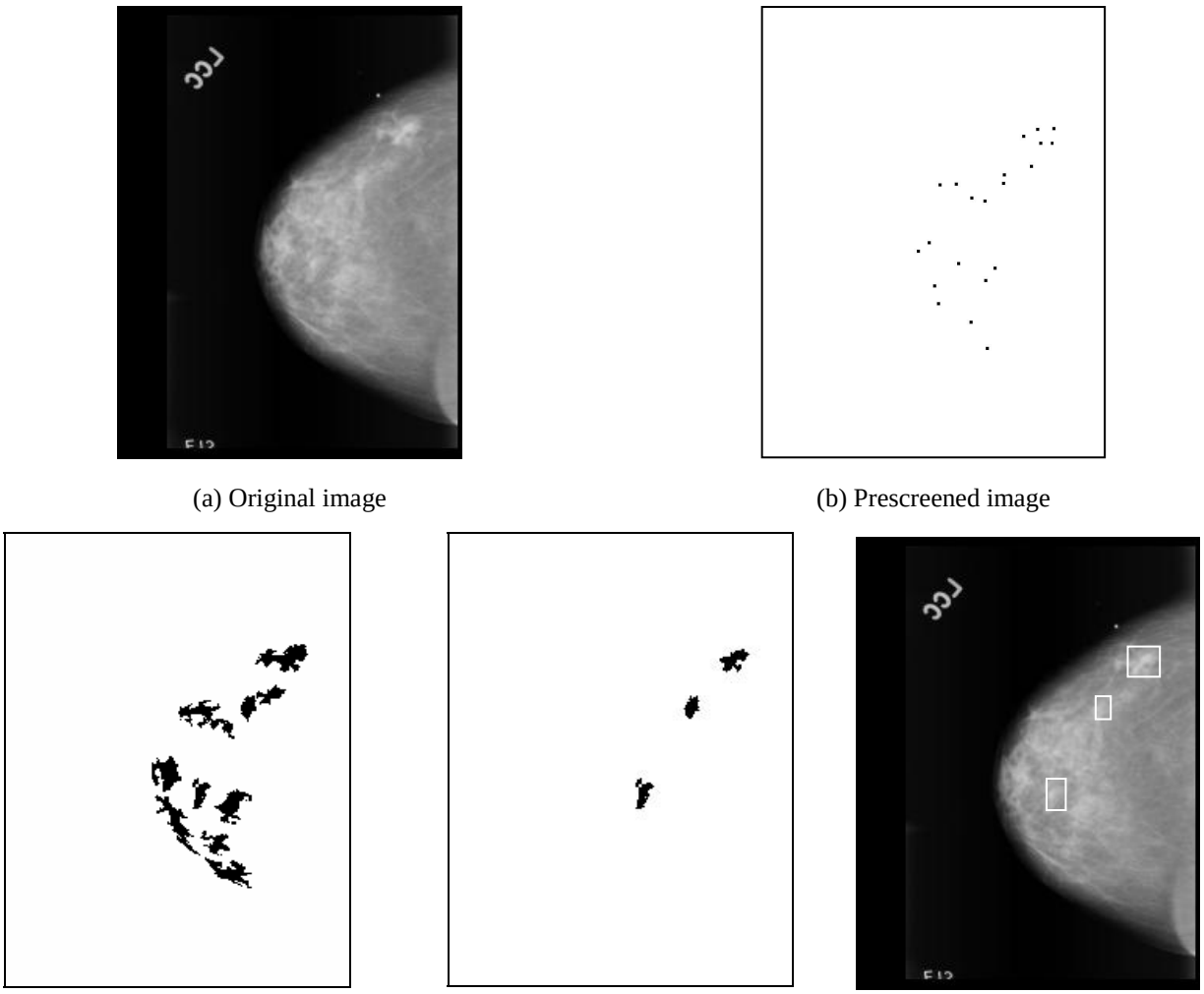


Figure 1. Block diagram of a single CAD system for mass detection on mammograms.

Our single-view CAD system consists of five processing steps: 1) pre-screening of mass candidates, 2) identification of suspicious objects, 3) extraction of morphological and texture features, and 4) classification between the normal and the abnormal regions by using rule-based and LDA classifiers. The block diagram for the single-view CAD system is shown in Figure 1. Figure 2 shows an example demonstrating the processing steps with our computer-aided mass detection system. For the pre-screening stage, we have developed a two-stage gradient field analysis method which combines the shape information of masses on mammograms with the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis. The gradient field analysis is used to determine locations of high convergence of radial gradient in the image. A region of interest (ROI) is then identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method in which the location of high gradient convergence is used as the starting point. Figures 2(b) and 2(c) show the initial detection locations and the grown objects, respectively. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. The objects whose new gradient convergence is lower than 80% of the original value are rejected. After prescreening, the suspicious objects are identified by using a clustering-based region growing method. For each suspicious object, eleven morphological features are extracted. Rule-based and LDA classifiers are trained to remove the detected normal structures that are substantially different from breast masses. Global and local multiresolution texture analyses are performed in each ROI by using the spatial gray level dependence

(SGLD) matrices at different pixel spacings and angular directions. In order to obtain the best feature subset and reduce the dimensionality of the feature space to design a robust classifier, feature selection with stepwise linear discriminant analysis is performed. Finally, LDA classification is used to identify potential breast masses. Figure 2(d) shows the final detected objects, and Figure 2(e) shows the locations of these objects superimposed on the mammogram.



(a) Original image (b) Prescreened image

(c) Identified suspicious objects (d) Detection result (e) Image with detected objects

Figure 2. An example demonstrating the processing steps with our single-view CAD system for mass detection.

2.2.2 Two-View Fusion

In order to improve the overall performance of our CAD system for detection of masses, we developed a two-view fusion technique which combines the information from two mammographic views. The fusion method used in this study is based on the assumption that the corresponding true mass on two different mammographic views will exhibit similarities in their geometric, morphological and textural features which are relatively invariant with respect to the imaging views. On the other hand, FPs detected by CAD system are expected to exhibit a lesser degree of similarity because they are usually objects formed by different normal tissues.

For a given object on one view, geometric pairing is first performed using the nipple-to-object distance as the average radius of an annular region on the other view within which the detected objects can be paired with the given

object. Manually identified nipple locations are used for the registration in this study. We are developing an automated nipple detection technique[26] and the automated method will be used when it reaches high accuracy. Similarity measures between each pair of objects are derived from the pairs of individual object features. The similarity features include morphological features, Hessian feature, correlation coefficients between the two paired objects and texture features. A similarity classifier is trained to distinguish between true and false pairs by merging the similarity features into a similarity score for each object. The similarity score and the single-view object score of the object are then fused to form a final score for the object. Our two-view system is summarized in Figure 3.

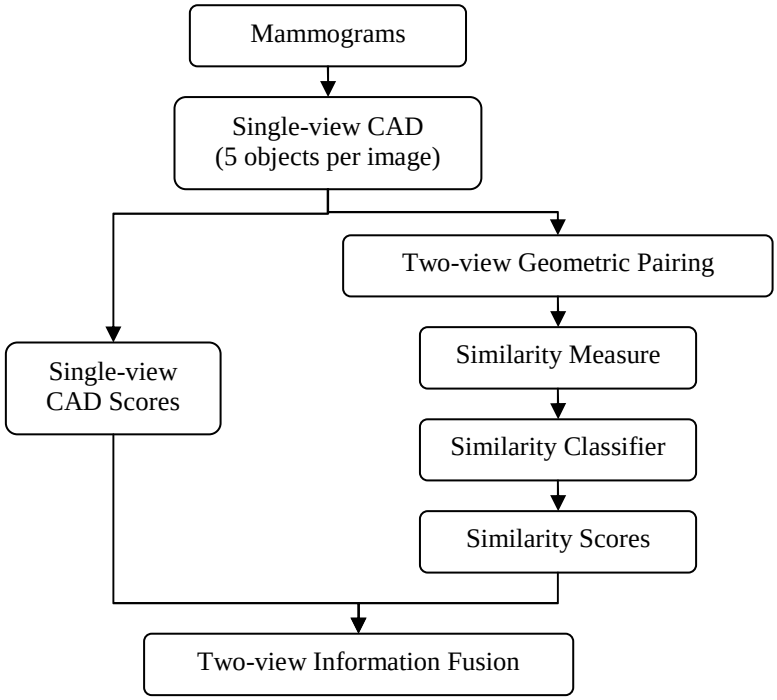


Figure 3. Block diagram of the two-view CAD system for mass detection on mammograms.

3. Experimental Results

We randomly separated the cases in our data set into two independent equal sized data sets: 243 cases with 494 images and 232 cases with 478 images. The training and testing were performed using the 2-fold cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and the detection of either or both masses on the two views was considered to be a true-positive (TP). To evaluate the overall test performance, an average test FROC curve was obtained from averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. When the single-view CAD system was applied to the test set, the FPs/image were 2.0, 1.5, and 1.2 at the case-based sensitivities of 90%, 85% and 80%, respectively. With the two-view CAD system, the FP rates were improved to 1.7, 1.3, and 1.0 FPs/image at the same case-based sensitivities. Figure 4 and 5 shows the comparison of the test performance of the single-view CAD system and the two-view CAD systems by using image-based and case-based average FROC curves, respectively. To analyze the improvement in the FROC curves statistically, an alternative free-response ROC (AFROC) method[27] was employed. In the AFROC method, false-positive images (FPI) instead of FPs per image are counted. The confidence rating of an FPI is determined by the highest confidence FP decision on the image regardless of how many lower confidence FP decisions

are made on the same image. The ROCKIT software developed by Metz et al[28] is used to analyze the AFROC data. The comparison of the A_1 and the p values is summarized in Table 1.

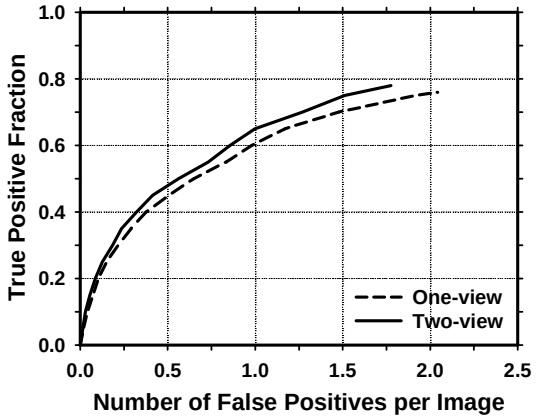


Figure 4. Image-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single-view: detection by the single-view CAD system. Two-view: detection by the two-view CAD system.

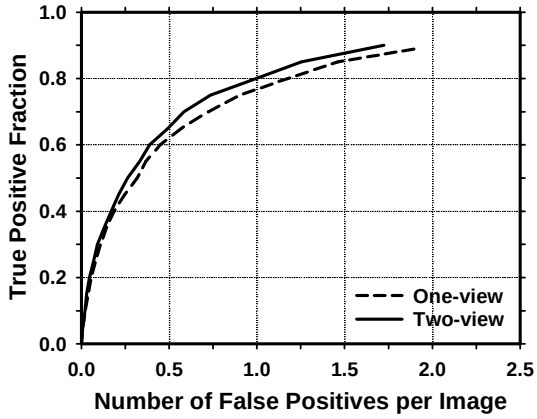


Figure 5. Case-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single-view: detection by the single-view CAD system. Two-view: detection by the two-view CAD system.

Table 1. Estimation of the statistical significance in the difference between the FROC performances of the single-view CAD system and the two-view CAD system.

	A_1 (AFROC)	
	Test Set 1	Test Set 2
One-view CAD	0.52	0.51
Two-view CAD	0.55	0.54
P Value	<0.0001	<0.0001

4. DISCUSSION AND CONCLUSIONS

In this study, we developed a two-view CAD system to improve the computerized detection of masses on mammograms. The two-view CAD system is different from case-based scoring, in which detection of the same mass in either the CC view or the MLO view will be counted as a true positive, in that the detected objects in the two views are correlated and analyzed for similarity and the likelihood score of a mass detected in both views may be enhanced compared with FPs. Our results indicate that two-view fusion can significantly improve the overall performance of the single-view CAD system. Future work will include automated identification of nipple locations and optimization of the fusion scheme in our system.

ACKNOWLEDGMENTS

This work is supported by USPHS grant CA95153, U. S. Army Medical Research and Materiel Command grants DAMD 17-02-1-0214 and W81XWH-04-1-0475. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC and ROCKIT programs.

REFERENCES

1. "American cancer society, www.Cancer.Org 2004, "Statistics for 2004"," American Cancer Society (2004).
2. C. R. Smart, R. E. Hendrick, J. H. Rutledge and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
3. C. Byrne, C. R. Smart, C. Cherk and W. H. Hartmann, "Survival advantage differences by age: Evaluation of the extended follow-up of the breast cancer detection demonstration project," *Cancer* **74**, 301-310, 1994.
4. S. A. Feig and R. E. Hendrick, "Risk, benefit, and controversies in mammographic screening," *Syllabus: A categorical course in physics technical aspects of breast imaging*, A. G. Haus and M. J. Yaffe, Eds., 119-135, Radiological Society of North America, Inc, 1993.
5. H. Seidman, S. K. Gelb, E. Silverberg, N. Laverda and J. A. Lubera, "Survival experience in the breast cancer detection demonstration project," *CA Cancer J Clin.* **37**, 258-290, 1987.
6. B. J. Hillman, L. L. Fajardo, T. B. Hunter, B. Mockbee, C. E. Cook, R. M. Hagaman, J. C. Bjelland, C. S. Frey and C. J. Harris, "Mammogram interpretation by physician assistants," *Am J Roentgenology* **149**, 907-911, 1987.
7. L. W. Bassett, D. H. Bunnell, R. Jahanshahi, R. H. Gold, R. D. Arndt and J. Linsman, "Breast cancer detection: One versus two views," *Radiology* **165**, 95-97, 1987.
8. M. G. Wallis, M. T. Walsh and J. R. Lee, "A review of false negative mammography in a symptomatic population," *Clinical Radiology* **44**, 13-15, 1991.
9. J. A. Harvey, L. L. Fajardo and C. A. Innis, "Previous mammograms in patients with impalpable breast carcinomas: Retrospective vs blinded interpretation," *American Journal of Roentgenology* **161**, 1167-1172, 1993.
10. R. E. Bird, T. W. Wallace and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613-617, 1992.
11. C. Beam, P. Layde and D. Sullivan, "Variability in the interpretation of screening mammograms by us radiologists," *Arch Intern Med* **156**, 209-213, 1996.
12. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
13. M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, C. Blane, D. Hoff, K. Hunt, L. Joynt, K. Klein, C. Paramagul, S. Patterson and M. A. Roubidoux, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - a pilot clinical trial," *Radiology* **231**, 208-214, 2004.
14. T. E. Cupples, "Impact of computer-aided detection (CAD) in a regional screening mammography program," *Radiology* **221(P)**, 520, 2001.
15. R. L. Birdwell, P. Bandodkar and D. M. Ikeda, "Computer-aided detection with screening mammography in a university hospital setting," *Radiology* **236**, 451-457, 2005.
16. S. L. Ng and W. F. Bischof, "Automated detection and classification of breast tumors," *Computers and Biomedical Research* **25**, 218-237, 1992.
17. N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," *Medical Physics* **23**, 1685-1696, 1996.
18. B. Zheng, Y. H. Chang and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," *Academic Radiology* **2**, 959-966, 1995.
19. N. Karssemeijer and G. Te Brake, "Detection of stellate distortions in mammograms," *IEEE Transactions on Medical Imaging* **15**, 611-619, 1996.
20. H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Trans Med Img* **15**, 235-245, 1996.
21. W. Qian, L. H. Li and L. P. Clarke, "Image feature extraction for mass detection in digital mammography: Influence of wavelet analysis," *Medical Physics* **26**, 402-408, 1999.
22. N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients.," *Radiology* **224**, 217-224, 2002.

23. S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner and M. A. Helvie, "Improvement of computerized mass detection on mammograms: Fusion of two-view information," *Medical Physics* **29**, 238-247, 2002.
24. B. Zheng, W. F. Good, D. R. Armfield, C. Cohen, T. Hertzberg, J. H. Sumkin and D. Gur, "Performance change of mammographic CAD schemes optimized with most-recent and prior image databases," *Academic Radiology* **10**, 283-288, 2003.
25. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Medical Physics* **32**, 2827-2838, 2005.
26. C. Zhou, H. P. Chan, C. Paramagul, M. A. Roubidoux, B. Sahiner, L. M. Hadjiiski and N. Petrick, "Computerized nipple identification for multiple image analysis in computer-aided diagnosis," *Medical Physics* **31**, 2871-2882, 2004.
27. D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Med Phys* **16**, 561-568, 1989.
28. C. E. Metz, "Roc methodology in radiologic imaging," *Investigative Radiology* **21**, 720-733, 1986.

Regularized discriminant analysis for breast mass detection on full field digital mammograms

Jun Wei, Berkman Sahiner, Yiheng Zhang, Heang-Ping Chan, Lubomir M. Hadjiiski,
Chuan Zhou, Jun Ge, Yi-Ta Wu
Department of Radiology, University of Michigan, Ann Arbor

ABSTRACT

In computer-aided detection (CAD) applications, an important step is to design a classifier for the differentiation of the abnormal from the normal structures. We have previously developed a stepwise linear discriminant analysis (LDA) method with simplex optimization for this purpose. In this study, our goal was to investigate the performance of a regularized discriminant analysis (RDA) classifier in combination with a feature selection method for classification of the masses and normal tissues detected on full field digital mammograms (FFDM). The feature selection scheme combined a forward stepwise feature selection process and a backward stepwise feature elimination process to obtain the best feature subset. An RDA classifier and an LDA classifier in combination with this new feature selection method were compared to an LDA classifier with stepwise feature selection. A data set of 130 patients containing 260 mammograms with 130 biopsy-proven masses was used. All cases had two mammographic views. The true locations of the masses were identified by experienced radiologists. To evaluate the performance of the classifiers, we randomly divided the data set into two independent sets of approximately equal size for training and testing. The training and testing were performed using the 2-fold cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. The average test FROC curve was obtained by averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. At the case-based sensitivities of 90%, 80% and 70% on the test set, our RDA classifier with the new feature selection scheme achieved an FP rate of 1.8, 1.1, and 0.6 FPs/image, respectively, compared to 2.1, 1.4, and 0.8 FPs/image with stepwise LDA with simplex optimization**.** Our results indicate that RDA in combination with the sequential forward inclusion-backward elimination feature selection method can improve the performance of mass detection on mammograms. Further work is underway to optimize the feature selection and classification scheme and to evaluate if this approach can be generalized to other CAD classification tasks.

Keywords: computer-aided detection, full field digital mammogram, mass detection, regularized discriminant analysis, feature selection

1. INTRODUCTION

Breast cancer is the most common cancer among American women[1]. Early detection and diagnosis can significantly increase the survival rate[2-4]. Recent clinical studies have shown that computer-aided detection (CAD) systems are helpful for increasing radiologists' accuracy in detecting breast cancers[5-8].

We have been developing CAD systems for detection and characterization of mammographic masses and microcalcifications. Detection of masses on mammograms is more challenging than detection of microcalcifications because the normal fibroglandular tissue in the breast causes false positives (FPs) by mimicking masses and causes false negatives due to overlapping with the lesions. Therefore, mass detection systems generally have lower sensitivity and higher FP rate than microcalcification detection systems. We are investigating methods to improve the overall performance of our CAD systems.

False positive (FP) classification is an important step in a CAD system. The basic approach in two-class classification is to assign an unknown sample to one of the two classes on the basis of a multidimensional feature space. A number of methods have been proposed in previous studies[9-11]. Most of the methods are based on linear discriminant analysis (LDA), artificial neural networks, and rule-based classifiers[12]. Recently, support vector machines were used to

Medical Imaging 2006: Image Processing, edited by Joseph M. Reinhardt, Josien P. W. Pluim,
Proc. of SPIE Vol. 6144, 61445P, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.654897

classify the malignant and benign clustered microcalcifications on mammograms[13]. In medical imaging application, a main problem during the classifier design is the finite sample size available which biases the performance of the trained classifier for unknown cases. In this study, we are investigating the performance of a regularized discriminant analysis (RDA) classifier in combination with a feature selection method for classification of the masses and normal tissues detected on full field digital mammograms (FFDMs).

2. MATERIALS AND METHODS

2.1 Materials

IRB approval was obtained prior to the commencement of this investigation. The images used in this study were acquired at the University of Michigan with a GE Senographe 2000D FFDM system before biopsy. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of $100\mu m \times 100\mu m$ and 14 bits per pixel. A data set of 130 cases was used. All cases had two mammographic views, the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral (LM or ML) view. The data set contained 130 biopsy-proven masses. The true locations of the masses were identified by a Mammography Quality Standards Act radiologist.

2.2 Methods

2.2.1 Discriminant Analysis

Assume that the class distributions are multivariate normal in a two-class classification problem. Under this condition, discriminant analysis models differ essentially by the specific assumptions on the mean vectors and covariance matrices of the group conditional densities. The most commonly used model is linear discriminant analysis (LDA) which assumes that the group conditional distributions are multivariate normal distributions with mean vectors μ_k, where k = 1, 2 is the class index, and equal covariance matrix Σ. The definition of LDA is given in Eq. (1).

$$Y = (\mu_1 - \mu_2)^T \Sigma^{-1} X \tag{1}$$

where $X^T=(x_1, \ldots, x_n)$ is the feature vector of a sample and n is the dimensionality of the feature space. If the covariance matrices are not equal, one can use quadratic discriminant analysis (QDA), which has a quadratic term for the feature vector in its model. The definition of QDA is described in Eq. (2).

$$Y = \frac{1}{2} X^T (\Sigma_2^{-1} - \Sigma_1^{-1}) X - (\mu_2^T \Sigma_2^{-1} - \mu_1^T \Sigma_1^{-1}) X \tag{2}$$

The parameters in LDA and QDA are usually unknown and have to be estimated from training samples. In medical imaging applications, the sample size may be very small in comparison with the dimensionality of the feature space. A regularization technique for discriminant analysis, referred to as regularized discriminant analysis (RDA)[14], makes use of a complexity parameter and a shrinkage parameter to design an intermediate classification model between LDA and QDA. The covariance matrices can thus be written as:

$$\hat{\Sigma}_k = (1-\gamma)\Sigma_k + \frac{\gamma}{p} tr[\Sigma_k] I \quad , \qquad k=1, 2 \tag{3}$$

where I is the identity matrix, γ and p are the complexity parameter and the shrinkage parameter, respectively. In this work, we investigated the use of the RDA classifier for FP reduction in a mass CAD system.

2.2.2 CAD System Overview

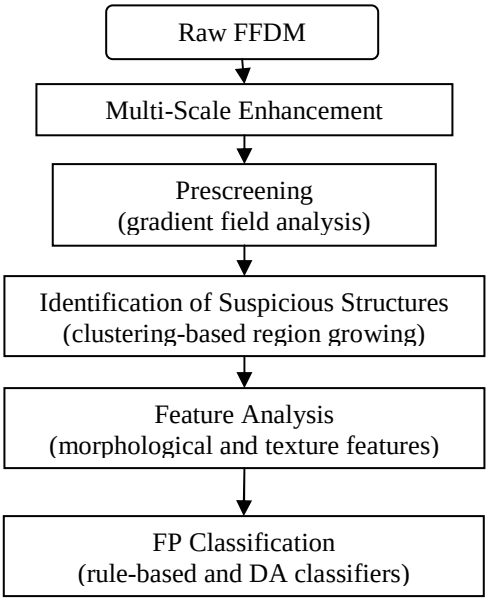
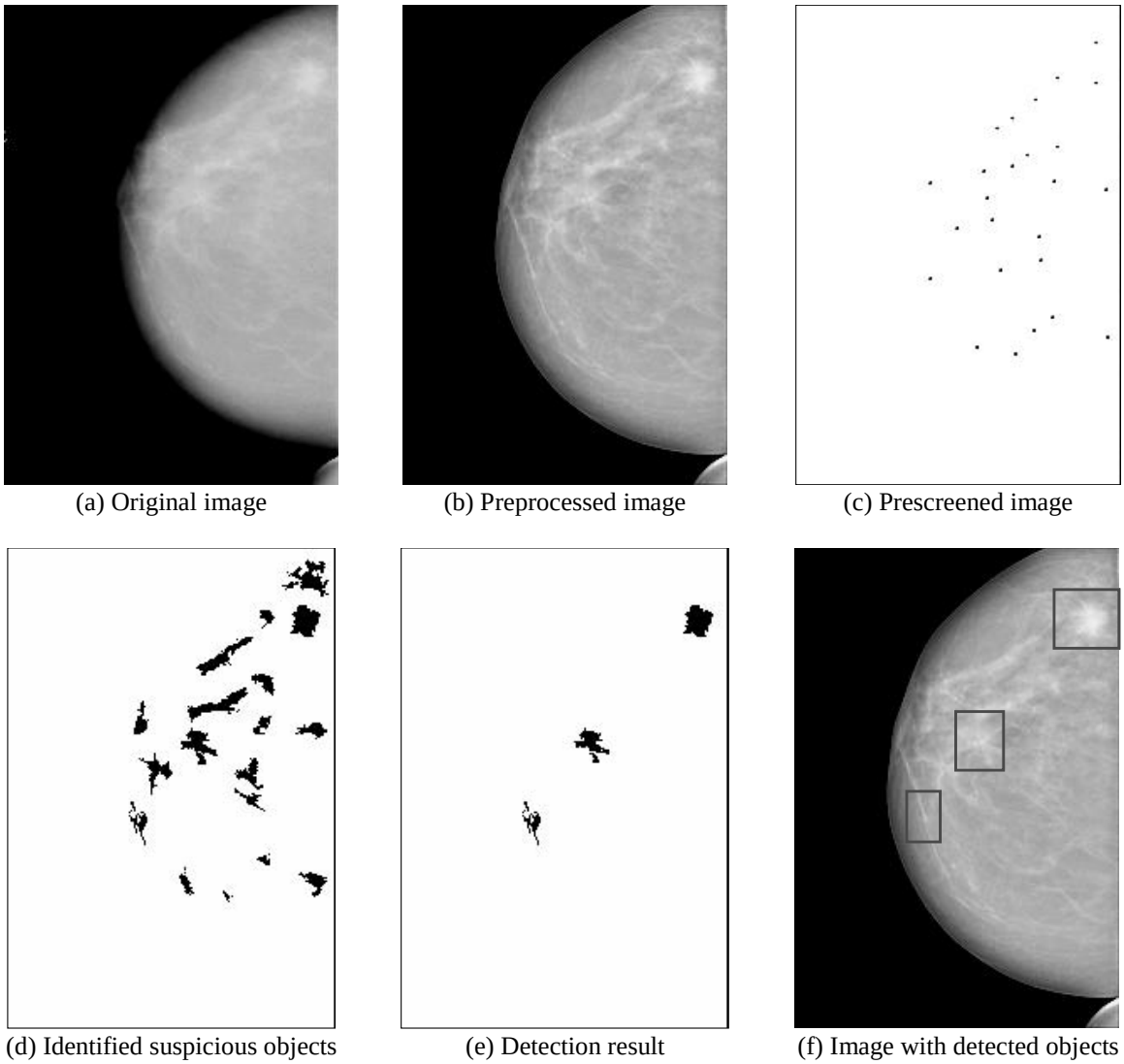


Figure 1. Block diagram of CAD system for mass detection on FFDMs.

Our CAD system consists of five processing steps: (1) preprocessing by using multi-scale enhancement, (2) pre-screening of mass candidates, (3) identification of suspicious objects, (4) feature extraction and analysis, and (5) FP reduction by classification of normal tissue structures and masses. The block diagram for the scheme is shown in Figure 1. FFDMs generally are pre-processed with proprietary methods before being displayed to readers. In an effort to develop a CAD system that is less dependent on specific FFDM systems, the raw digital images are used as input to our system. A preprocessing scheme based on a multi-resolution method[15] has been developed for image enhancement. This scheme consists of three steps. First, the boundary of the breast is detected automatically by using Otsu's method[16]. Second, the Laplacian pyramid is used to decompose the image into multi-scales. A nonlinear weight function is designed to enhance each high-pass component. Finally, the Gaussian pyramid is used to reconstruct the multi-scales. An example of an original mammogram and the enhanced mammogram are shown in Figs. 2(a) and 2(b), respectively. After preprocessing, gradient field analysis was used to detect the mass candidates from the preprocessed FFDMs. The suspicious objects are then identified by using a clustering based region growing method. Figures 2(c) and 2(d) show the initial detection locations and the grown objects, respectively. For each suspicious object, eleven morphologic features are extracted and rule-based and discriminant classifiers are trained to remove the detected normal structures that are substantially different from breast masses. Global and local multiresolution texture analysis[17, 18] are performed in each region of interest by using the spatial gray level dependence matrix. Finally, discriminant classification is used to identify potential breast masses. Further details of this algorithm can be found in the literature[19].

In order to obtain the best texture feature subset and reduce the dimensionality of the feature space to design an effective classifier, feature selection was applied to the training set. Stepwise LDA feature selection with Wilks' lambda as the selection criterion was employed in our previous study. Simplex optimization procedure was used to choose the best set of feature selection parameters which includes a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for excluding features that have high correlation with the features already in the selected pool. In this study, we compared a new stepwise feature selection procedure with the current method. In the proposed method, a feature selection scheme which combines forward stepwise feature selection and backward stepwise feature elimination is used to obtain the best feature subset, using the area under the receiver operating characteristic (ROC) curve, A_z, as the selection criterion instead of Wilks' lambda. We evaluated the classifier performance using a leave-one-case-out resampling scheme within the training set, the test discriminant scores from the left-out cases were analyzed using ROC methodology. The discriminant scores were input as the decision variable in

the LABROC program, which fits a binormal ROC curve based on maximum likelihood estimation. The performances of the RDA classifier and the LDA classifier, both with the new feature selection method, were compared to that of the LDA classifier using the Wilks' lambda as the stepwise feature selection criterion in terms of their A_z for the classification of masses and normal tissue.



(a) Original image (b) Preprocessed image (c) Prescreened image

(d) Identified suspicious objects (e) Detection result (f) Image with detected objects

Figure 2: An example demonstrating the processing steps with our computer-aided mass detection system.

3. RESULTS

We randomly separated the cases in our data set into two independent data subsets: 66 and 64 cases. The training and testing were performed using the cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered as an independent true object. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and the detection of either or both masses on the two views was considered to be a true-positive (TP). The average test FROC curve was obtained by averaging the FP rates at the same

sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. The CAD system using RDA with the new feature selection method achieved an image-based sensitivity of 60%, 65%, and 70% at 1.1, 1.4, and 1.6 FPs/image, respectively, compared with 1.4, 1.7, and 2.1 FPs/image for the CAD system using LDA with the new feature selection method. The CAD system with stepwise LDA and simplex optimization achieved FP rates of 1.6, 1.9, and 2.2 FPs/image, respectively, at the same sensitivities, which were comparable to the FP rates of the CAD system using LDA with the new feature selection method. For case-based FROC analysis, the results are summarized in Table 1. Figures 3 and 4 show the comparison of the image-based and case-based average FROC curves of the CAD systems using the three different classification methods, respectively.

Table 1. Comparison of case-based performance of three methods. OFS: stepwise feature selection with simplex optimization. NFS: feature selection combining forward feature selection and backward feature elimination.

TP	FPs/image		
	LDA-OFS	LDA-NFS	RDA-NFS
70%	0.8	0.7	0.6
80%	1.4	1.3	1.1
90%	2.1	2.2	1.8

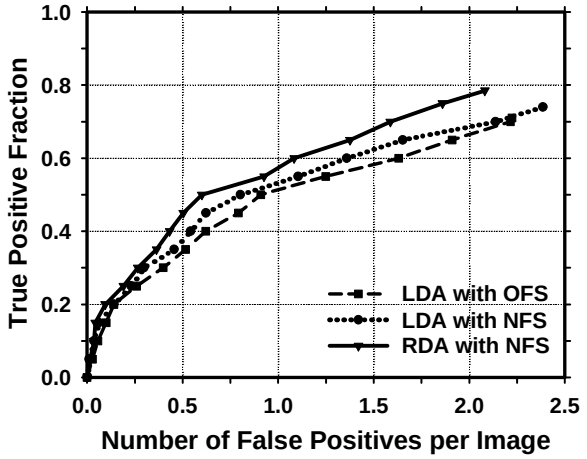


Figure 3. Comparison of image-based FROC curves. OFS: stepwise feature selection with simplex optimization. NFS: feature selection combining forward feature selection and backward feature elimination.

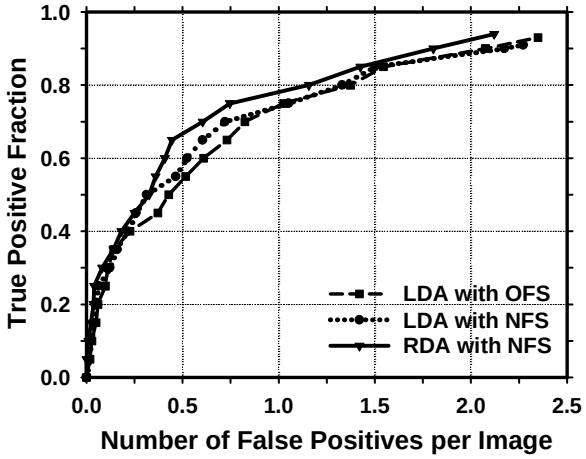


Figure 4. Comparison of case-based FROC curves. OFS: stepwise feature selection with simplex optimization. NFS: feature selection combining forward feature selection and backward feature elimination.

4. DISCUSSION AND CONCLUSIONS

We previously developed a CAD system for detection of masses on FFDMs. In this study, we investigated the use of an RDA classifier with a new feature selection method. Our results indicated that the new FP classifier can improve the overall performance of our CAD system. Further work is underway to optimize the feature selection and classification scheme and to evaluate if this approach can be generalized to other CAD classification tasks.

ACKNOWLEGMENTS

This work is supported by U. S. Army Medical Research and Materiel Command grants DAMD 17-02-1-0214 and W81XWH-04-1-0475 and USPHS grant CA95153. The content of this paper does not necessarily reflect the

position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

REFERENCES

1. "American cancer society, www.Cancer.Org 2004, "Statistics for 2004"," American Cancer Society (2004).
2. C. R. Smart, R. E. Hendrick, J. H. Rutledge and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
3. S. A. Feig, C. J. D'orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger and P. Wilcox-Buchalla, "American college of radiology guidelines for breast cancer screening," *AJR Am J Roentgenol.* **171**, 29-33, 1998.
4. B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," *CANCER* **91**, 1699-1703, 2001.
5. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
6. S. V. Destounis, P. Dinitto, W. Logan-Young, E. Bonaccio, M. L. Zuley and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," *Radiology* **232**, 578-584, 2004.
7. M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, C. Blane, D. Hoff, K. Hunt, L. Joynt, K. Klein, C. Paramagul, S. Patterson and M. A. Roubidoux, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - a pilot clinical trial," *Radiology* **231**, 208-214, 2004.
8. R. L. Birdwell, P. Bandodkar and D. M. Ikeda, "Computer-aided detection with screening mammography in a university hospital setting," *Radiology* **236**, 451-457, 2005.
9. H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Medical Physics* **22**, 1555-1567, 1995.
10. H. P. Chan, B. Sahiner, R. F. Wagner and N. Petrick, "Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers," *Medical Physics* **26**, 2654-2668, 1999.
11. B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler and M. M. Goodsitt, "Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images," *IEEE Transactions on Medical Imaging* **15**, 598-610, 1996.
12. Q. Li and K. Doi, "Analysis and minimization of overtraining effect in rule-based classifiers for computer-aided diagnosis," *Med. Phys* **33**, 320–328, 2006.
13. J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Med.Phys* 2005 (Submitted).
14. J. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association* **84**, 165-175, 1989.
15. P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications* **COM-31**, 337-345, 1983.
16. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. System, Man, Cybernetics* **9**, 62-66, 1979.
17. D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograpms: Multiresolution texture analysis," *Medical Physics* **22**, 1501-1513, 1995.
18. D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," *Medical Physics* **24**, 903-914, 1997.
19. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Medical Physics* **32**, 2827-2838, 2005.

Computer aided detection of breast masses on prior mammograms

Jun Wei*, Berkman Sahiner, Heang-Ping Chan, Lubomir M. Hadjiiski, Marilyn A. Roubidoux, Mark A. Helvie, Jun Ge, Chuan Zhou, Yi-Ta Wu
Department of Radiology, The University of Michigan, Ann Arbor, MI 48109

ABSTRACT

An important purpose of a CAD system is that it can serve as a second reader to alert radiologists to subtle cancers that may be overlooked. In this study, we are developing new computer vision techniques to improve the detection performance for subtle masses on prior mammograms. A data set of 159 patients containing 318 current mammograms and 402 prior mammograms was collected. A new technique combining gradient field analysis with Hessian analysis was developed to prescreen for mass candidates. A suspicious structure in each identified location was initially segmented by seed-based region growing and then refined by using an active contour method. Morphological, gray level histogram and run-length statistics features were extracted. Rule-based and LDA classifiers were trained to differentiate masses from normal tissues. We randomly divided the data set into two independent sets; one set of 78 cases for training and the other set of 81 cases for testing. With our previous CAD system, the case-based sensitivities on prior mammograms were 63%, 48% and 32% at 2, 1 and 0.5 FPs/image, respectively. With the new CAD system, the case-based sensitivities were improved to 74%, 56% and 35%, respectively, at the same FP rates. The difference in the FROC curves was statistically significant ($p<0.05$ by AFROC analysis). The performances of the two systems for detection of masses on current mammograms were comparable. The results indicated that the new CAD system can improve the detection performance for subtle masses without a trade-off in detection of average masses.

Keywords: computer-aided detection, prior mammogram, mass detection, AFROC analysis

1. INTRODUCTION

Breast cancer is one of the leading causes of cancer mortality among women[1]. Studies indicate that radiologists do not detect all carcinomas that are visible upon retrospective analyses of the images[2-8]. Computer-aided diagnosis (CAD) is considered to be one of the promising approaches that may improve the sensitivity of mammography[9, 10].

An important application of a CAD system is to serve as a second reader to alert radiologists to subtle cancers that may be overlooked. Masses retrospectively seen on prior mammograms represent the difficult cases that are more likely to be missed by radiologists. To study the ability of a CAD system in detecting subtle cancers, one way is to evaluate its accuracy in detecting missed cancers on prior mammograms. Our previous experiences indicate that CAD schemes trained with cancers on current images do not perform well in detecting masses seen retrospectively on prior images[11]. In this study, we designed new techniques to improve the detection performance for subtle masses on prior mammograms and also evaluated the new CAD system on both prior and current mammograms by comparing with our previously developed CAD system[12].

2. MATERIALS AND METHODS

2.1 Materials

All mammograms in this study were collected from patient files in the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. The mammograms were digitized with a LUMISYS 85 laser film scanner with a pixel size of 50μm×50μm and 4096 gray levels. The scanner was calibrated to have a linear relationship between gray levels and optical densities (O.D.) from 0.1 to greater than 3 O.D. units. The nominal O.D.

* jvwei@umich.edu, phone: 734-647-8553, CGC B2103, 1500 E. Medical Center Dr., Ann Arbor, MI 48109-0904

Medical Imaging 2007: Computer-Aided Diagnosis, edited by Maryellen L. Giger, Nico Karssemeijer,
Proc. of SPIE Vol. 6514, 651405, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.713768

range of the scanner is 0–4. The full resolution mammograms were first smoothed with a 2×2 box filter and subsampled by a factor of 2, resulting in images with a pixel size of 100μm×100μm. These images were used for the input of our CAD system. The data set we used in this study contained 159 patients. Each exam had two mammographic views, resulting in a total of 318 current mammograms and 402 prior mammograms. Forty-two patients had two years of prior examinations. All mammograms were obtained before biopsy. There were 159 biopsy-proven masses in this data set. Figures 1 and 2 showed the histograms of mass sizes and visibility, respectively, for the comparison of current and prior masses. The size of a mass was estimated as its longest diameter seen on the mammograms. The visibility of the masses was rated by an experienced radiologist on a 10-point scale with 1 representing the most visible masses and 10 the most difficult case relative to the cases seen in clinical practice. The mass size ranged from 3 to 42 mm (mean size: 14.3±8.6 mm on current mammograms and 10.9±6.6 mm on prior mammograms) and the visibility ratings extended over the entire range. For the current mammograms, 140 of the masses were visible on both views and 19 visible on only one view. For the prior mammograms, 100 masses were visible on both views and 101 visible only on one view. Therefore, there were 299 visible and 19 invisible masses on current mammograms and 301 visible and 101 invisible masses on prior mammograms if the masses were counted independently by mammographic view.

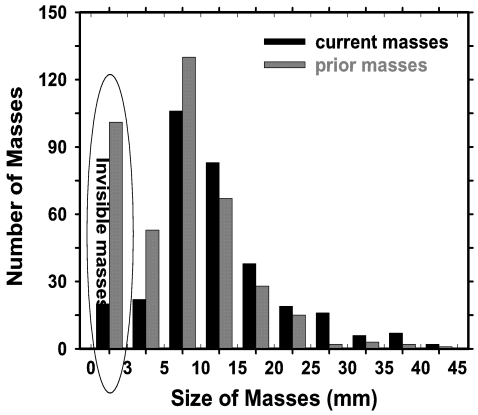


Figure 1. Histogram of the sizes for 299 masses on current mammograms and 301 masses on prior in our data set. Mass sizes are measured as the longest dimension of the mass by an experienced MQSA radiologist. The size of the masses in this data set ranged from 3 to 42 mm.

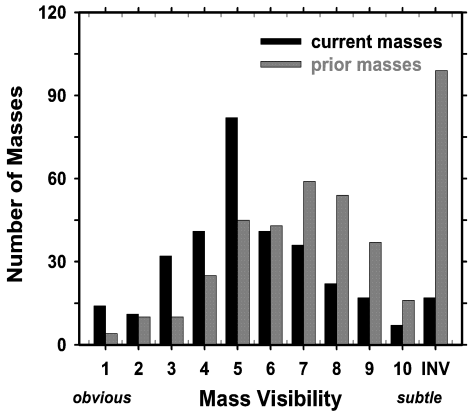


Figure 2. Histogram of the visibility of the masses in our data set. The visibility is evaluated on a 10-point rating scale with 1 representing the most visible masses and 10 the most difficult case relative to the cases seen in their clinical practice. The masses that were not visible were plotted in the column labeled as "INV".

2.2 Methods

2.2.1 CAD System Overview

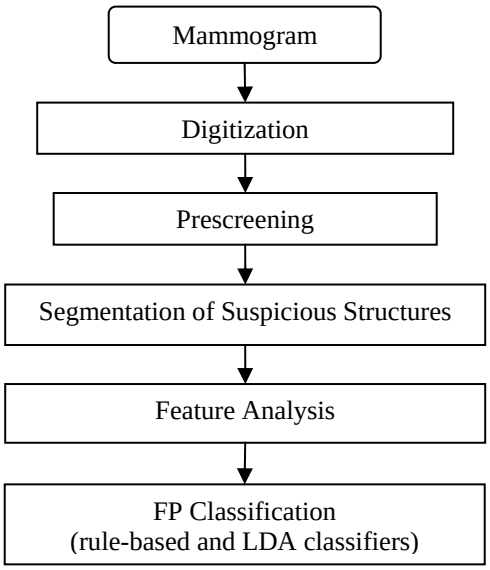


Figure 3. Block diagram of a single CAD system for mass detection on mammograms.

Our CAD system consists of five processing steps: 1) pre-screening of mass candidates, 2) identification of suspicious objects, 3) extraction of morphological and texture features, and 4) classification between the normal and the abnormal regions by using rule-based and LDA classifiers. The block diagram for the CAD system is shown in Figure 3.

For the pre-screening stage, we developed a new prescreening technique in which gradient field analysis was combined with Hessian analysis to identify mass candidates. Both gradient field and Hessian analyses were designed to enhance circular structures on mammograms and to suppress the objects with other shapes. Gradient field analysis used the information of gradient field directions and Hessian analysis used the second derivatives by solving for the eigenvalues of the Hessian matrix. After this enhancement filtering, the local maxima within the breast region were identified as the mass candidates on each mammogram. The suspicious structure in each identified location was initially extracted by a seed-based region growing method. An active contour method was then used to further refine the initial segmentation. Morphological, gray level histogram and run-length statistics (RLS) features were extracted from the original region of interest (ROI) and the orientation field of the ROI for reduction of FPs.

2.2.2 Training and test CAD system

The hold-out method was used for training and testing our CAD system. We randomly separated the entire data set by case into two independent subsets, the training subset including 78 cases with 156 current and 200 prior mammograms and the test subset including 81 cases with 162 current and 202 prior mammograms. The training included selection of proper parameters and features for the classifier in the CAD system. Once the training was completed, the parameters and features were fixed for testing. The new system was trained by using prior mammograms in the training set only. The performance of the new system was compared with that of the previous CAD system on the current and prior mammograms in the test set.

During training, feature selection with stepwise LDA was employed to obtain the best feature subset and reduce the dimensionality of the feature space to design an effective classifier. The detailed procedure has been described

elsewhere[13]. Briefly, at each step one feature was entered or removed from the feature pool by analyzing its effect on the selection criterion, which was chosen to be the Wilks' lambda in this study. Since the appropriate threshold values for feature entry, feature elimination, and tolerance of feature correlation were unknown, we used an automated simplex optimization method to search for the best combination of thresholds in the parameter space. The simplex algorithm used a leave-one-case-out resampling method within the training subset to select features and estimate the weights for the LDA classifier. To have a figure-of-merit to guide feature selection, the test discriminant scores from the left-out cases were analyzed using receiver operating characteristic (ROC) methodology. The accuracy for classification of masses and FPs was evaluated as the area under the ROC curve, A_z. In this approach, feature selection was performed without the left-out case so that the test performance would be less optimistically biased. However, the selected feature set in each leave-one-case-out cycle could be slightly different because every cycle had one training case different from the other cycles. In order to obtain a single trained classifier to apply to the hold-out test subset, a final stepwise feature selection was performed with the best combination of thresholds, found in the simplex optimization procedure, on the entire training subset to obtain the final set of features and estimate the weights of the LDA. Note that the entire process of feature selection and classifier weight estimation was performed within the training subset. The LDA classifier with the selected feature set was then fixed and applied to the test subset.

2.2.3 Evaluation methods

We used a free-response receiver operating characteristic (FROC) method to assess the overall performance of the CAD scheme on this image set. An FROC curve was obtained by plotting the mass detection sensitivity as a function of FP marks per image as the decision threshold on the LDA classifier scores varied. The detected individual objects were compared with the "true" mass locations marked by the experienced radiologist, as described above. A detected object was labeled as TP if the overlap between the bounding box of the detected object and the bounding box of the true mass relative to the larger of the two bounding boxes was over 25%. Otherwise, it would be labeled as FP. The 25% threshold was selected as described in our previous study[14].

FROC curves were presented on a per-image and a per-case basis. For image-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to the number of visible masses by image, which was 149 and 151, respectively, for the current and prior test subset. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either or both masses on the two views was considered to be a TP detection; the sensitivity was thus calculated relative to the number of masses by case, which was 81 and 90, respectively, for the current and prior test subset. The test FROC curve for a given mass subset was estimated by counting the detected masses on the test mass subset for the sensitivity. The FP marker rate was estimated from FPs detected in the same test subset. The average number of FP marks per image produced by the CAD system at a given sensitivity was estimated by counting the detected objects in these cases at the corresponding decision threshold.

In order to compare the performance of our CAD systems statistically, we employed the alternative free-response ROC (AFROC) method[15]. In the AFROC method, the FROC data are first transformed by counting the number of false-positive images (FPI) instead of the FPs per image. The LDA score of an FPI is determined by the FP object with the highest score on the image regardless of how many lower scores FP objects are made on the same image. The ROCKIT curve fitting software and statistical significance tests for ROC analysis developed by Metz et al. [16] can then be used to analyze the AFROC data.

3. EXPERIMENTAL RESULTS

Figures 4 and 5 showed the image-based and case-based FROC curves for detection of masses on prior mammograms, respectively. The case-based sensitivities for detection of masses on the prior mammograms (typically subtle masses) in the test subset were 56%, and 35% at 1 and 0.5 FPs/image by using the new CAD system in comparison to 48%, and 32% at the same FP rates by using the previous CAD system. The improvement with the new system on prior mammograms was statistically significant ($p = 0.036$). When the new system was applied to the detection of masses

on the current mammograms (typically average masses) in the test subset, the case-based sensitivities were 77% and 70% at 1 and 0.5 FPs/image in comparison to 75% and 56% at the same FP rates by using the previous CAD system. The difference in the two FROC curves for detection of average masses on current mammograms was not statistically different ($p = 0.184$). Image-based and case-based FROC curves for detection of masses on current mammograms were shown in Figures 6 and 7, respectively.

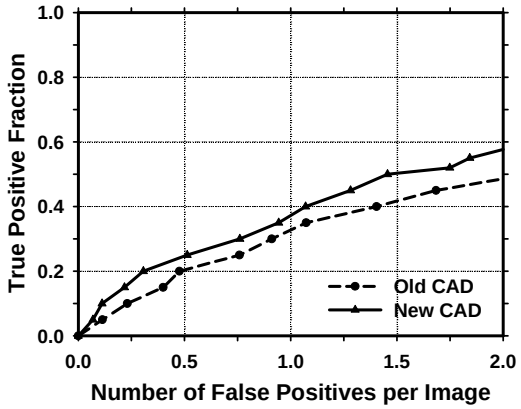


Figure 4. Image-based test FROC curves on prior mammograms. Old CAD: detection by the previous CAD system trained on both current and prior mammograms. New CAD: detection by the CAD system trained on prior mammograms.

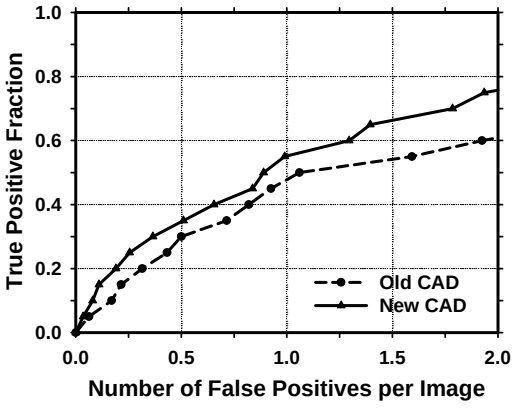


Figure 5. Case-based test FROC curves on prior mammograms. Old CAD: detection by the previous CAD system trained on both current and prior mammograms. New CAD: detection by the CAD system trained on prior mammograms.

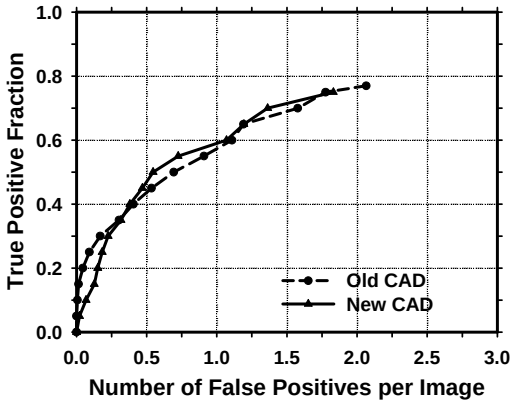


Figure 6. Image-based test FROC curves on current mammograms. Old CAD: detection by the previous CAD system trained on both current and prior mammograms. New CAD: detection by the CAD system trained on prior mammograms.

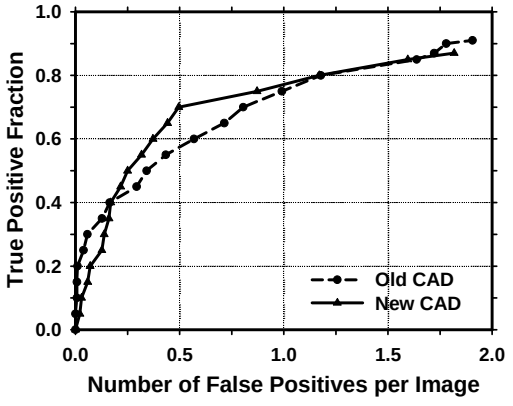


Figure 7. Case-based test FROC curves on current mammograms. Old CAD: detection by the previous CAD system trained on both current and prior mammograms. New CAD: detection by the CAD system trained on prior mammograms.

Table 1. Estimation of the statistical significance in the difference between the FROC performances of the previous CAD system trained on both current and prior mammograms and the proposed CAD system trained on prior mammograms.

	A_1 (AFROC)	
	Current Test Set	Prior Test Set
Old CAD	0.51	0.26
New CAD	0.50	0.31
p-value	0.184	0.036

4. DISCUSSION AND CONCLUSIONS

In this study, we improved the accuracy of a CAD system for detection of subtle masses on prior mammograms. A new prescreening method was developed to improve the sensitivity of mass detection. A new mass segmentation method that combined a seed-based region growing method with active contour method was also designed. RLS features were extracted from the original ROIs and the newly derived orientation field of the ROIs for FPs reduction. Our CAD system can significantly improve the performance of mass detection on prior mammograms without a trade-off in the detection of masses on current mammograms. It is expected that the new CAD system can increase the overall accuracy for detection of subtle early-stage breast cancers.

ACKNOWLEGMENTS

This work is supported by USPHS grant CA95153, U. S. Army Medical Research and Materiel Command grants DAMD 17-02-1-0214 and W81XWH-04-1-0475. The content of this paper does not necessarily reflect the position of the funding agencies and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC and ROCKIT programs.

REFERENCES

1. "Cancer statistics 2006 presentation, www.Cancer.Org," American Cancer Society (2006).
2. B. J. Hillman, L. L. Fajardo, T. B. Hunter, B. Mockbee, C. E. Cook, R. M. Hagaman, J. C. Bjelland, C. S. Frey and C. J. Harris, "Mammogram interpretation by physician assistants," *AJR* **149**, 907-911, 1987.
3. L. W. Bassett, D. H. Bunnell, R. Jahanshahi, R. H. Gold, R. D. Arndt and J. Linsman, "Breast cancer detection: One versus two views," *Radiology* **165**, 95-97, 1987.
4. M. G. Wallis, M. T. Walsh and J. R. Lee, "A review of false negative mammography in a symptomatic population," *Clinical Radiology* **44**, 13-15, 1991.
5. J. A. Harvey, L. L. Fajardo and C. A. Innis, "Previous mammograms in patients with impalpable breast carcinomas: Retrospective vs blinded interpretation," *AJR* **161**, 1167-1172, 1993.
6. R. E. Bird, T. W. Wallace and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613-617, 1992.
7. C. A. Beam, P. M. Layde and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists - findings from a national sample," *Archives of Internal Medicine* **156**, 209-213, 1996.
8. V. Beam, D. Sullivan and P. Layde, "Effect of human variability on independent double reading in screening mammography," *Academic Radiology* **3**, 891-897, 1996.
9. F. Shtern, C. Stelling, B. Goldberg and R. Hawkins, "Novel technologies in breast imaging: National cancer institute perspective," *2nd Postgraduate Course Syllabus*. 153-156, 1995.
10. C. J. Vyborny, "Can computers help radiologists read mammograms?," *Radiology* **191**, 315-317, 1994.
11. J. Wei, H.-P. Chan, B. Sahiner, L. M. Hadjiiski, M. A. Helvie, M. A. Roubidoux, C. Zhou and J. Ge, "Dual system approach to computer-aided detection of breast masses on mammograms," *Medical Physics* **33**, 4157-4168, 2006.
12. J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, M. A. Helvie, M. A. Roubidoux, C. Zhou, J. Ge and Y. Zhang, "Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms," *Proc. SPIE* **6144**, 241-247, 2006.

13. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Medical Physics* **32**, 2827-2838, 2005.
14. N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault and L. M. Hadjiiski, "Breast cancer detection: Evaluation of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients.," *Radiology* **224**, 217-224, 2002.
15. D. P. Chakraborty and L. H. L. Winter, "Free-response methodology: Alternate analysis and a new observer-performance experiment," *Radiology* **174**, 873-881, 1990.
16. C. E. Metz, "ROC methodology in radiologic imaging," *Investigative Radiology* **21**, 720-733, 1986.